Familiarity effects in visual word recognition

Possidonia de Freitas Drumond Gontijo

Degree of Doctor of
Philosophy
University of Edinburgh
1998

To
Iris , Salu
&
Ivair

Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself with the exception of the programming part contained in Chapter 7, that was done in collaboration with Ivair Gontijo.

Signed:

Date: ...

Abstract

This thesis is an investigation of two different aspects of familiarity processes involved in visual word recognition. The first is how capitalisation influences visual word recognition. The second is the role played by onset, nucleus and coda in nonword recognition.

A familiar aspect of proper names in English, is that they are printed with an initial capital letter. Two experiments investigated the effects of the capitalisation of the initial letter of nonwords. It was found that subjects generate fewer pronunciations for initially capitalised nonwords than for those which were not capitalised. I suggest that in English initial capitalisation acts as a cue strong enough to prompt readers to perceive unfamiliar strings of letters as belonging to the category of proper names. As a result, the phonological domain used to retrieve the pronunciation of initially capitalised strings becomes more restricted than that used for the non-capitalised unfamiliar strings. These results extend the applicability of Brennen's theory for proper names, which is based on the size of the set of plausible phonologies of a word.

In a third experiment, pairs of nonwords had their familiar visual appearance manipulated in terms of first and last letter capitalisation, in a same-different matching task. Faster response times were obtained for those nonword pairs that kept a more familiar aspect (e.g., pairs in which the first letter was capitalised as opposed to others in which the last letter was capitalised). These results are explained in terms of Besner and Jonhston (1989) "orthographic familiarity route". I propose the *transformation model* as an explanation for the mechanisms by which this route operates.

Nonwords are an important aspect of this thesis. A new algorithm was developed for the creation of monosyllabic nonwords in which the frequency of their onsets, nuclei and codas could be controlled carefully. This gave us the opportunity to study the influence of orthographic neighborhood in visual word recognition. The findings here are in agreement with previous studies which show the recognition of an item to be influenced by the presence of neighbours.

It has been hypothesized that familiarity effects in visual word recognition can only be found in tasks where identification mechanisms are not implicated. Here, a new category of words, namely brand names, was used to test this hypothesis. There are many reasons why brand names are a more appropriate class of words than acronyms to be used in this type of investigation. The results obtained confirm the hypothesis above. Previously, acronyms had been the only class of words used to test this hypothesis.

Finally, a computational assessment of the nature of the mappings from letter-to-sound in British English was carried on. A program was developed to estimate the pronunciation of any string of English graphemes based on the probabilities of grapheme-phoneme correspondences. The algorithm was assessed by examining its behavior for nonwords. This was done by using a corpus of nonword transcriptions, collected in an experiment with trained phoneticians. The results confirm the fact that the statistical information about grapheme-phoneme correspondences alone is not sufficient to predict English pronunciation. Also, a method was developed that allows the quantification of the different orthographic depth for various languages.

Acknowledgments

Many people laid the foundation for this thesis. A process that started quite a long time ago when at a very early age I learnt the importance of reading. Here, I wish to express my thanks to the many people responsible for that. In particular, my grandmother whom at a time when books were difficult to acquire gave me many of hers, including the seventeen volumes of " The Arabian Nights".

Much later in life, after a first degree and an M.Phil. in Philosophy, I became a PhD student in the Centre for Cognitive Science. Here, I was led by the enthusiasm and knowledge of my first supervisor, Richard Shillcock, into investigating some of the fascinating mechanisms that make it possible for humans to acquire reading skills. Later, Richard went away to Australia in a sabbatical leave, and I was informed that Louise Kelly, as my second supervisor would take charge of the supervision. During this time we worked perfectly in tune and her contribution to my PhD has been very important. I thank both of them most sincerely for all their kindness, prompt help and forbearance in my moments of stricken panic.

As I found out, writing a thesis can produce enjoyable moments, but it is also an exhausting and wearily enterprise. My gratitude extends too all my colleagues here in CogSci, deserving special mention: Andrew Gillies, Cathy Sotillo, Daniel Liu, Holly Branigan, Karen Budewig and Padraic Monaghan, Renata Vieira and Saturnino Luz for making my life interesting and cheerful.

I am especially grateful for Merce Prat's friendship. Having started our PhD together we shared the delights of academic discussions and also supported each other at all moments, the good and the bad, throughout our lives as PhD students.

Betty and Dyane, with their efficiency and kindness have always succeeded in making me feel confortable and at home in CogSci.

My thanks also go to CST (Computer Support Team), for their efforts in overcoming the difficulties I found in this initially foreigner environment.

Irene MacCleod deserves a big thanks for having opened to me the doors of the Speech Laboratory in the Linguistic Department. Also, for having taught me the beginnings of how to make the best use of the *Waves+* program.

The Brazilian Research Council CAPES- Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior, generously supported this research and I am pleased to acknowledge them.

The support of my family and Ivair's family has also been important to me in this long journey. I wish to thank them unreservedly for that. Joaquim Martins, my uncle, who have supported me intelectually and financially during my early years as an undergraduate student, has given me the means without which I could never have achieved what I did. To him my heartly felt thanks.

Finally, this thesis is dedicated to my parents with my deepest regrets that they did not live to see it. It is also dedicated to Ivair, my husband, who faithfully believed in me, and gave me the strength for winning the small and the big obstacles I found standing in my path.

CONTENTS

...He was a very enthusiastic and fairly good pupil and would be absolutely amazed at his own progress. Sometimes, during a lesson, he would suddenly get up, take a book from the shelf, raise his eyebrows high and read two or three lines after great effort. His face would turn red and he would look at me and say in an astonished voice: "You see, I can read, ever hear anything like it!" Then he would close his eyes and repeat some poetry:

"Just like a mother mourning over the grave of her son,
So sails the sandpiper over the desolate plain..."

Read it?

Several times he would ask cautiously, almost in a whisper:
"Tell me, my friend, how it all comes about. A man looks at these commas and hyphens and they turn into words and I recognise them, they're our living words! How do I know this? No one's whispering them into my ear. If these were pictures, then I could understand. But here it seems that the thoughts themselves are printed on the page - how do they do it?
What could I answer? My "don't know" annoyed him. "The work of a magician!" he said sighing, as he peered at the pages and held them up to the light...

(Maxim Gorky teaching a peasant how to read. Extract from "My universities", pp 111, 1979, London: Penguin Press).

Chapter 1

1.1 - Introduction

This thesis is about familiarity effects in the visual perception of strings of letters. In this area studies are traditionally carried on using displays consisting of single stimulus object. Therefore, this thesis is also about visual word recognition.

Visual word recognition has proved to be one of the most popular topic of study among researchers interested in the theme of cognition (Besner & Humphreys, 1995). This popularity emerges from scientific as much as from sociological and pragmatic factors. The simplicity involved in the presentation of isolated strings of letters on a computer screen, combined with the degree of sensitivity and accuracy of the experimental techniques used in visual word recognition, have conspired to make it attractive to researchers. Another enticement is the ready access to procedures for online measurement of reaction time, with precision of the order of one millisecond. This enables the cognitive scientist to measure small and subtle effects with great accuracy. Also, the impressive performance of connectionist networks in simulating visual word recognition has opened up numerous new avenues of research interest.

The bedrock for scientists working in visual word recognition is their trenchant belief that word processing serves as the basis for more complex processes. Also, important is the belief that visual word recognition can operate independently of the other cognitive modules and in this way it presents a unique opportunity for studying a specific process in isolation from other factors.

The primary aspect of word recognition that this thesis concentrates on is visual familiarity. Although it is generally recognised that the cultural familiarity of a linguistic display affects visual information processing, the specific processes that underlie familiarity effects are not yet fully understood. The traditional problem of whether familiarity with a stimulus affects processes concerned with the extraction and encoding of its physical aspects is still being studied. As will be discussed later, the finding of an adequate solution to this issue has subtle but important implications for those concerned with a more accurate modelling of visual word recognition.

Next, an overview of the thesis structure is provided in the form of a brief description of each of its chapters.

1.2 - The structure of the thesis

Chapter 1 - Offers an overview of the thesis structure.

Chapter 2 - Presents a literature review of the critical issues dealt with in the thesis. The review is split into four major sections. The first is an introduction to some of the most influential models of visual word recognition. Next, different views of "what are the visual units of word recognition" are examined. The third section discusses the role of some variables that can be quantified at the word level in visual word recognition. Finally, a brief review of the literature concerning brand names and nonwords is given.

Chapter 3 - Deals with the methodological aspects of the experimental work. It describes and comments on the suitability of the types of experimental paradigms, variables and measurements that have been used throughout the thesis.

Chapter 4 - Discusses one of the possible subtleties involved in the reading process: the role played by initial capitalisation as a clue in the reading of proper names. I report two experiments in which it was found that subjects produced a smaller number of pronunciations for initially capitalised nonwords than for non-capitalised ones. These results are interpreted in the light of Brennen's set size plausible phonology theory. We discuss the implications of these issues in relation to connectionist models of pronunciation and suggest some adjustments to their architectures. We have also investigated previous findings that neighbourhood density is a determining factor in pronunciation by using nonwords that were manipulated according to a variable which was characterized as "weirdness". I find that larger number of different pronunciations were produced for nonwords with more sparse neighbourhoods (weird group) than for the nonweird group.

Chapter 5 - Considers the types of visual information that might plausibly be used to recognise strings of letters. We report a same-different matching-task experiment run with nonwords. The results show that, at least for tasks such as the same-different matching task paradigm, the more accustomed people are to certain physical patterns, the faster they process them. The orthographic familiarity route of Besner & Johnston (1989) is explored as an attempt to explain the experimental results. Some of the possible mechanisms that might be in place allowing for the use of the orthographic familiarity route are then proposed. Finally, a neuropsychological case is revisited in a PostScript to the chapter.

Chapter 6 - This chapter examines the hypothesis that familiarity effects in visual word recognition are affected by the nature of the task performed (Besner et. al., 1984), by extending previous work done with acronyms to a new category of words, i.e., brand names. We report two experiments; a naming task and a lexical decision task. While familiarity effects are found in the lexical decision task, no effect is found for the naming task. We discuss these results on the light of Besner and colleagues hypothesis.

Chapter 7 - More and more statistical approaches are being applied to the study of language. This chapter reports a computational assessment of the nature of the mappings from letter to sound in British English. This information is used to estimate the pronunciation of any string of English graphemes. The usefulness and limitations of this pronunciation prediction algorithm are discussed. The algorithm was further assessed by examining its behaviour for nonwords. A corpus of nonwords transcriptions was collected from a set of trained phoneticians: these provided the baseline for assessment of the algorithm. Finally, considerations are made with respect to the limitations of a solely statistical approach to the pronunciation of English.

Chapter 8 - In a general discussion I consider the implications of the thesis findings for models of word recognition, for current and future research. Finally, I present a summary of the thesis in terms of the questions it has answered.

Chapter 2

Issues in visual word recognition

How does the reader recognise a printed word? This question is asked time and time again in the psycholinguistics literature. Answering it would bring us a great deal closer to understanding the nature of mental representation. This thesis explores familiarity effects on the recognition and pronunciation of visually presented strings and the results are considered in terms of visual word recognition models.

Theoretical models play a fundamental role in the rational organisation of empirical evidence that otherwise would only be a disconnected bundle of information. Although there are currently several competing models of visual word recognition, in this chapter I examine the development of such models by looking at three which have had historical impact, namely, the search, the logogen and the interactive-activation models.

Next, the two rival theories of pattern recognition known as the template-matching and the feature-analysis theories will be discussed in terms of their contribution to the models of visual word recognition. This is followed by an overview and discussion of the different candidates to the role of functional units in visual word recognition. Some of the variables that can be quantified at the whole word level are also discussed here since some of them are relevant to the present work. Finally, an account is offered of the two stimuli that have been frequently used throughout the experimental work to be described here, i.e., nonwords and brand names.

2.1 - Models of word recognition

Models are the means by which new experimental evidence is accommodated in a scene of already established theoretical foundations. Thus, they represent a good parameter for measuring the scientific progress achieved in specific areas. In the field of visual word recognition the situation is not different, with old models being replaced by new ones and also with intense competition happening between those models that simultaneously share the burden of explanation.

In this section, we pay tribute to three earlier models of visual word recognition, namely, the search, the logogen and the interactive-activation models. They are discussed briefly in terms of those features that were a source of inspiration for models of a later generation. The concept of "logogen units", found in the interactive-activation model is a reverberation of Morton's logogen model. More recent models will be dealt with in later chapters to ground some of the explanations for our experimental results.

First, however, it will be introduced the concept of a *mental lexicon* and that of an *access code* for lexical access. I start by acknowledging that the form of the mental lexicon is disputed, with some researchers adopting a distributed and others a local conception of how information is to be represented in the mental lexicon. However, it is useful to examine the concept in its broadest definition. This is specially true for the models to be described next, since they are also known as models of lexical access.

To read a word is to extract information from a set of printed marks and then to use that information as a means of reaching the word's lexical entries. This process is lexical access. The mental lexicon is envisaged as a repository of all the information a reader or listener has attained about words of his language (Coltheart, 1977; Treisman, 1960, 1961). From the point of view of recognising words visually, the mental lexicon can be described metaphorically as composed of at least three compartments, each one holding a different type of information that characterises a particular word. These compartments or lexical entries specify the meaning, the pronunciation and the spelling of each word. This idea will become clear from the description of our first model, i.e., the search model.

Among the numerous questions about the mental lexicon that have to be answered for the reading process to be understood there is one that, as we shall see later in Chapter 5 is of special interest to us here. It is concerned with the *access code* for the mental lexicon, that is, how does the reader proceed from the information he extracts from a word stimulus to the corresponding lexical entry? In other words, what is the nature of the information extracted from a printed word for use in lexical access? According to the phonological view, a printed word is first converted into a corresponding phonological representation, using a system of phoneme-grapheme correspondence rules. This phonological information is the *name code*, which is subsequently used to represent the printed word during the process of lexical access. For the proponents of the visual route, it is purely visual information from the printed word that is used during lexical access. This visual information might be for example, individual letter shapes, overall word-shape etc. The most common view adopted today, according to Rayner and Pollatsek (1990), is that virtually all lexical access during reading is done by the direct visual route.

To recap, the three models to be described next are precursors in visual word recognition. They are no longer valid per se, in the sense that they do not account for all the available data on the processing of word recognition. Our results will not be conducted strictly in terms of these models; we use them as simple tools for exploring some of the issues studied in word recognition.

2.1.1 - The search model

The search model is based upon the idea that, in response to input information, the lexicon is searched serially until a suitable match is found (Foster, 1976, 1989; Foster & Davies, 1984). However, for reasons of time and efficiency, it is very unlikely that the whole of the mental lexicon is scanned for a match every time a word is to be processed. A pure search model, in which the whole of the lexicon is searched every time a word is encountered, has never been seriously proposed. Foster suggested that each search is confined to a particular subset of the lexicon, which is defined on the basis of sensory characteristics. In the case of visual input, for example, defining a subset criterion could be: "all words starting with *t* and finishing with *h*". Another important characteristic of the items in a subset of the lexicon is that they are organised in terms of frequency, such that the more frequent items are searched first.

The model was designed to have two-stages in order to account for sensory input pertaining to different modalities. There are two types of files accessed during a search. At the first stage access is obtained through *access files*. The access files are modality-specific, i.e., there are different files for orthographic, phonological, and syntactic-semantic sources. These access files assign pointers to a *master file* in the lexicon which stores all information regarding a word, including its meaning. The architecture of the search model is depicted in Fig (2.1).

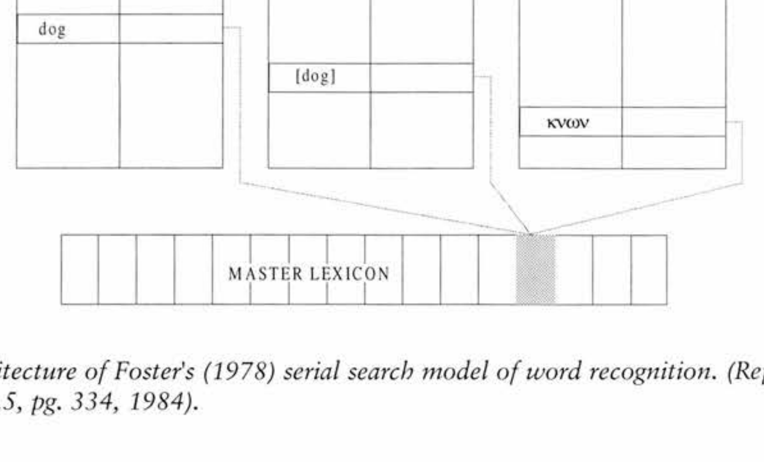

Fig. (2.1) - Architecture of Foster's (1978) serial search model of word recognition. (Reproduced from Balota, D.A, Fig.5, pg. 334, 1984).

Let us now take a visually presented string of letters and follow its processing through the search model. At the first level, the orthographic access file (or a subset of it) is serially searched and when a lexical representation is found that adequately matches the input letter string, this directly leads to the corresponding master file entry. The full spelling of the word can then be checked back to the original stimulus or its representation in memory. If they match, the letter string will be recognised as being that word. If they do not match, the search resumes in the orthographic access file in order to find a more suitable match.

The model accounts for the main data in word recognition, such as frequency and lexicality data. The grounds on which the plausibility of the serial search mechanism has been criticised is that other models not depending on serial search can account for the same data equally well. Additionally, Carr and Pollatsek (1985) show that serial search models have problems accounting for the recognition of long and morphologically complex words. Finally, the model does not convincingly account for how we process and pronounce nonwords, since nonwords can only be rejected by the model at the expense of an exhaustive searching process that does not seem to reflect empirical findings.

2.1.2 - The logogen model

One of the inspirations for this model has come from the Pandemonium model (Selfridge, 1959) to be described later. It can be said to be the ancestor of the connectionist models of visual word recognition; a fact that will become clear below. The model is illustrated in Fig. (2.2).

The architecture of the logogen system (Morton, 1969, 1979) is very similar to that of the search model in that both have separate orthographic and phonological input systems. The mechanism by which an entry is accessed however, is quite different as it will be seen shortly. Also, access to the lexical entries in the search model is done indirectly via the visual or auditory input. By contrast, in the logogen model, perceptual information feeds directly into the logogens, which are devices that collect evidence to be used in the process of revealing the identity of an appropriate word target. They can also be thought of as the mental representations of lexical items in this theory and are sensitive to information that may disclose the appropriateness of the target word. The evidence gathered by the logogens can be either perceptual or contextual, i.e., there is no distinction between evidence for a word from external and internal sources. For example, whenever low-level visual analysis identifies the feature "horizontal bar on top of the word", the feature count of all the logogens for words beginning with E F I J T Z is enhanced, if the word is in capital letters. Each logogen has a resting level energy of activation. As logogens receive corroborating evidence that corresponds to the stimulus presented, its activation levels increases. If this level increases beyond a threshold value, the logogen activates.

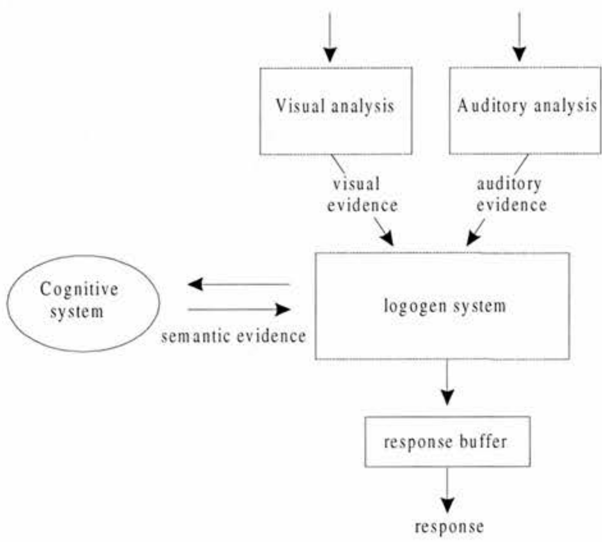

Fig. (2.2) - The main components and relationships of the logogen model. (Reproduced from Garman, 1990, Fig. 5.9, pg. 278).

The threshold is set so that it is reached only if it is almost certain that the logogen really corresponds to the input word. When the logogen is activated, the corresponding word is recognised and the logogen system prepares for the next input. All logogens are simultaneously active in collecting their specific information, i.e., the logogen system is a parallel accessing device. The information that activates logogens originates from the so-called *cognitive system*, which is the repository of all conceptual, syntactic and higher-order functions. The cognitive system replaces, in the logogen model, the *master file* we have seen in the search model.

Any model that tries to explain the mechanisms underlying word recognition has to account for the effects of frequency. In the present model this is done by assuming that logogens have different threshold levels with more frequent items having a lower threshold than the less frequent ones. Thus, the more frequent a word is, the more often it is encountered by the cognitive system and the lower the threshold of its corresponding logogen becomes.

The logogen model also accounts well for much of the basic data in word recognition. The problem posed by nonwords to the search model, for example, is dealt with here by having the system set a deadline for reaching a threshold. If no logogen has reached threshold by the deadline, the response is "no". However, as Coltheart et. al. (1977) have argued, this deadline should vary during the identification of a word. It starts at a low value, but is increased as the activation in the logogen system increases; if there is a lot of activityover few logogens, the stimulus is probably a word, and a "no" response should be made with caution.

2.1.3 - The interactive-activation model

As mentioned in the description of the previous model, the interactive-activation framework is an elaboration on the logogen approach, involving a level of lexical units that behave in a somewhat similar way to logogens. This is one of earliest connectionist models of high level processing and its architecture is shown in schematic form in Fig. (2.3).

The model was built with the assumption that perceptual processes take place within a system in which there are several hierarchical levels of processing that occur in parallel. It is an interactive network involving the activation of three input layers representing three processing levels: the feature level, the letter-level and the word level. At the word level, its architecture also allows for several letters of a word to be processed at one time. The basic idea is that when a word is presented to the cognitive system, its letter features activate consistent features at the feature level. These in turn activate the representations of letters that they are part of and inhibit inconsistent ones. At the same time, activated word representations feed back to reinforce activation of consistent and inhibit inconsistent letter representations. In addition, there are also within-layer inhibitory links, so that an active representation at one level will inhibit other representations at the same level. With time, as more features are extracted, activation accumulates for letters and words consistent with the representations. Gradually the system converges to a single interpretation of the input and word recognition takes place.

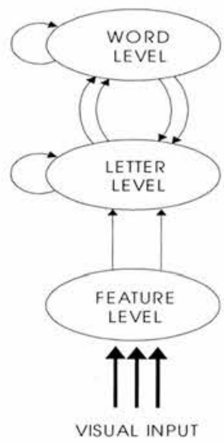

Fig.(2.3) - The interactive activation model of word recognition (Reproduced from McClelland & Rumelhart, 1981).

In the interactive-activation model, the word superiority effect is easily accounted for by the feedback mechanism between the different layers. As letters in the context of a word receive activations from the word units above them, they become easier to see in the context of a word than when presented in isolation. The original purpose of the interactive-activation model was to account for the word superiority effect and from this point of view it was very successful. The word superiority effect was originally described by Reicher (1969) and, independently, by Wheeler (1970). In short, it can be defined as the phenomenon by which letters are easier to identify, when rapidly presented in the context of a word, than when they are seen in isolation. An important consequence of having demonstrated the word superiority effect by means of the interactive-activation model is that the functional units of this model are therefore the individual letters and not the whole word. We will return to this point later in Chapter 5 where this subject is discussed more thoroughly.

2.1.4 - Summary

In this section we have called attention to the importance of scientific models in research and three of the most influential models in visual word recognition have been briefly described. The first model to be presented was search model, in which the lexicon is searched serially and indirectly after the stimuli has passed through the access files. The presentation of the logogen model followed. There the access to the lexicon is direct and processed in parallel by making use of features and word detectors called logogens. Finally, a description was given of the connectionist interactive-activation model that inherited part of its architecture from the logogen system.

We also flagged the Besner and McCann (1987) paper which is discussed further in Chapter 4. It is through models of this type that we will discuss the issues surrounding familiarity effects in visual word recognition.

2.2 - Pattern recognition

The previous section was a brief overview of models of visual word recognition. As we saw, all those models work with the assumption that the first step towards recognition is the encoding of primitive features by the cognitive system. The goal of the present section is to describe how this consensus has been brought about.

Pattern recognition is the process by which the brain extracts relevant information from a background of spurious signals and places an appropriate interpretation on the signals it has extracted. It can also be understood as the mapping of a specific pattern onto a more general pattern, i.e., the identification of a set of features as belonging to an object. In order to trust a pattern recognition system, its performance must be consistent, repeatable and reliable. Next, we offer a description of how visual lexical encoding is accomplished by means of two theories that are based in pattern recognition.

2.2.1 - The Template-matching theory

The simplest approach to pattern recognition involves template-matching. These templates are variously known as *prototypes* or *canonical forms*. It is assumed that there is a template for each of the patterns to be recognised and recognition is accomplished when an arriving input matches exactly an internal template. In the case of visual word recognition inputs letter-features could be letters or words. This is the case in the Fig. (2.4)-(a), where the letter *A* was used as an illustration.

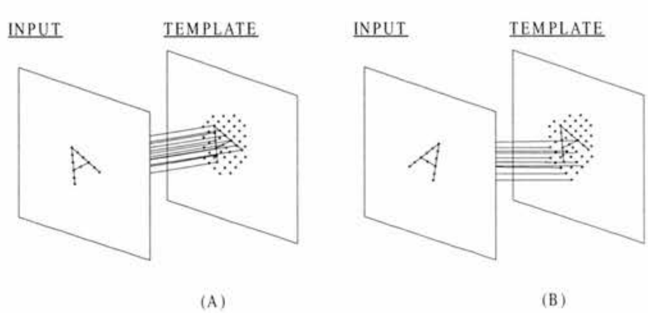

Fig.(2.4) - The template-matching system character and the rotational difficulties. (Based on Rayner & Pollatsek, 1980, pg.12).

To account for human pattern recognition performance however, the template-matching theory would have be much more flexible than the simple version presented above. Generally, humans do not have any problems in recognising, for example, forms that have gone through some transformations in their size, rotation or even that have missing parts. The template-matching device however can not account for these transformations. As it can be seen from Fig. (2.4)-(b) the *A* in the picture has been rotated and as a result the template can not match it any more. A solution to this problem would be to postulate a new template for each altered shape of the object, but that would mean the postulation of an infinite number of templates. This is surely not the best way of describing pattern recognition.

Another argument used against the template theory is that patterns such as letters, seem to be defined by critical local features rather than global shape. The difference between *Q* and *O*, for example, is the presence or absence of the bottom bar, not the exact shape of the circular body of the letter.

Although it has been argued that, with modifications, the template theory could cope with the problems exposed above, the new concept of feature detection mechanisms was introduced. Behavioural and neurological evidence began to accumulate during the 50s and the 60s that favoured the new approach and led later to the development of some influential models in pattern recognition.

Gibson and Gibson (1955) argued that pattern recognition was done by feature level analysis. If primitive features could be identified, then one could have the building blocks of pattern recognition. Instead of having to match whole templates, perhaps objects are recognised by detecting features like lines, corners and curves. The question then arose as to how the feature detectors are combined so as to allow the recognition of patterns of features as objects. One suggestion was that feature detectors are arranged in a hierarchy which starts by detecting simple features and ends with the detection of complex patterns.

Before describing one of the most influential systems that used the idea of feature detection let us mention very briefly the neurological and behavioural evidence that supported this view.

2.2.2 - Behavioural and neurological evidence

One of major contributions in the neurological area came from the recordings of the activity of single fibbers in the optic nerves of unanesthetized frogs, while presenting various visual stimuli. It was successfully shown that the frog's fibbers were quite selective in the kind of stimuli that activated specific cells (receptive-fields) in its retina. Some fibres would respond, for example, to small dark objects entering cats' receptive-fields, but would not respond to a large moving object. These results were interpreted as clearly showing that complex features of the input, not simply its parts are abstracted very early by the visual system (Lettvin, Maturana, McCulloch & Pitts, 1959). Similar experiments have also been performed with cats (Hubel & Wiesel, 1959; 1962). By recording the activation of single cells at several levels of the visual system, from the retina up to the visual cortex, it was found that some cells are activated by different kinds of inputs in their receptive fields. Lower level cells respond to simple inputs like a stationary dot, a line or a moving dot. Cells higher up in the visual system are only activated by complex configurations such as lines at specific orientations.

Positron emission tomography has also been used to extend this work to humans (Petersen et. al., 1990). Significant blood flow changes were detected in specific areas of the striate cortex corresponding to feature-like detection systems in alert humans.

Behavioural evidence for the feature analysis approach has also been obtained. Confusion matrices indicated that letters which shared features were more likely to be confused in degraded perceptual conditions, compared to letters which did not share many features (Kinney et. al., 1966). In addition, visual search studies indicated that subjects were relatively faster to find a given target letter (e.g., Z) when it was embedded in a set of letters that did not share many features with the target (e.g. O, J, U, Z, D), compared to a set of letters that did share many features with the target (e.g. F, N, K, Z, X) (Neisser, 1967). More recently the same effect was demonstrated for a different type of stimuli, in a study reported by Polk and Farah (1994), in which Canadian postal workers who dealt extensively with postal codes consisting of mixed strings of letters and numbers (e.g., EH8 9LW) were found to have a reduced "pop-out" effect for letters embedded in short string of numbers, compared with matched controls.

2.2.3 - The pandemonium model

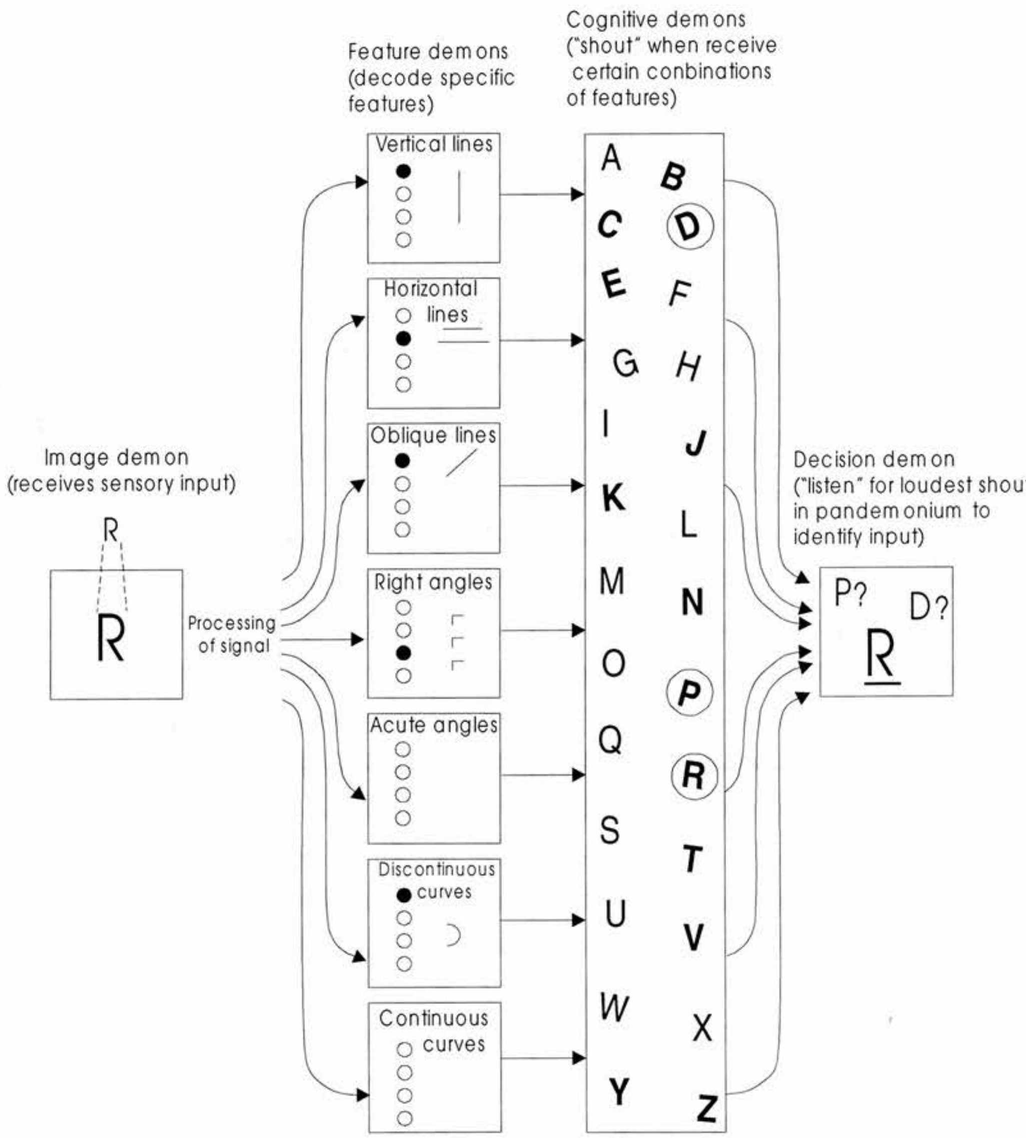

Fig. (2.5) - Selfridge & Neisser's (1960) Pandemonium model of letter recognition. (Reproduced from Cohen et. al., 1994)

One of the first well-specified feature analytic models of letter recognition was the *pandemonium model* (Selfridge, 1959). This explored the fact that letters are difficult to recognise due to the fact that different people write their letters differently.

The model's name - pandemonium - captures the fact that it is composed of *demons* that perform computations in parallel without attention to one another, and each *shouts out* its judgement of which letter had been presented. The fundamental assumption in the model is that the cognitive system used in searching is hierarchically organised. The pandemonium is organised such that it starts by detecting simple features (feature demons) and ends with the detection of complex patterns (decision demon). This is illustrated by Fig. (2.5) and these demons are each specialised in gathering evidence for one particular letter.

When a novel pattern is presented, its features are analysed and listed, and this input list is compared in parallel with each of the stored defining lists for each known pattern. The closer the input feature list matches one stored for a known pattern, the more *loudly* does the corresponding pattern demon *shout* and the pattern is classified as belonging to the category represented by the demon shouting the loudest. Also, the pandemonium system can learn to give more weight to some features than to others. This model can be implemented in terms of a network and one of its virtues is that it would still make a correct or plausible judgement about a letter even if some of its features were missing or atypical.

Not surprisingly, there are drawbacks to this model as well and we will mention two of them. However, both of them could perhaps be amended. One criticism is that the pandemonium model does not take context into account; yet humans are much more likely to decide that something is an O instead of a Q in the context of a three-letter word as "c-t". The other type of problem presented by the model lies not in its basic architecture, but in the nature of the representations it encodes. The features that have been considered in the development of such models are very local parts of patterns, like bars and angles. Feature lists of this kind are said not to be rigid enough, i.e., they do not contain all the necessary information for pattern recognition. For example, for something to be a letter T, the spatial arrangement of its features is important, not only the features themselves. The information that for something to be a "T" a vertical line must support and bisect the horizontal line is important. However, the relative lengths of the lines are not important. This is illustrated in Fig. (2.6).

Fig. (2.6) - Correct (a) and wrong (b) spatial arrangement of the letter "T". (Reproduction based on Bruce & Green, 1985).

To sum up, in this section a brief overview of the process of pattern recognition has been offered. Its relevance to present work is made clear when one considers that almost all models of visual word recognition has assumed feature analysis to be the first step into recognition.

2.3 - Sub-lexical levels of organisation

In this section we examine the different ways in which a word can be partitioned and the likelihood of each of these partitions to be seen as candidates to fill in the role of functional units of visual word recognition. There are four distinct levels of sublexical representations that could perform this role, namely, letters, onsets and rimes, morphology and syllables. However, only letters and onset/rimes will be discussed here, as only these two levels are directly relevant to this dissertation.

2.3.1 - Letters

The encoding of letter identities is seen by many as an important part, and quite possibly an essential precursor to the normal word reading process. It is therefore believed that the understanding of the mental operations and codes involved in the identification of letters is crucial for a detailed theory of reading. There is a vast literature on visual letter recognition, in which the subject is approached from many angles. There are psychophysical studies aiming at identifying, for example, the effects brought upon recognition by the manipulation of variables such as colour (Dumbar & McCleod, 1984; Dyer, 1971; Regan, 1978) rotation (Driver & Baylis, 1995; Friedman & Hall, 1996) and size (Brooks, 1977; Rudnicky & Kolers, 1984) of the stimuli. There are also studies that investigate the role of letters when these are part of a string context. Such strings are then controlled for variables such as length, orthographic familiarity, lexicality etc. Neuropsychological cases where brain damaged patients showing deficits in letter processing are also carefully investigated as a source of information for the role of letters in visual recognition (Patterson & Wilson, 1990; Arguin & Bub, 1992).

In more concrete terms, let us take one of the variables above as an example, say, the length of a word. The length of a word, quantified for instance, as the number of letters it possess, influences its reading. Studies have shown that these influences although present, depend on the type of task being performed by the subjects. The consequences of such findings have a direct impact on models of word recognition. They are problematic, for example, for Gough (1982) model of word recognition, where it is suggested that word recognition is processed letter by letter, in a serial fashion from left to right. If that is so strings containing larger number of letters would take longer to be processed independently of the type of task being performed. In fact, most psycholinguists (Rayner & Pollatsek, 1989) nowadays believe that the letters are processed not serially but in parallel by the cognitive system. Another important variable with regards to letter processing is that of the frequency with which a letter appears in print. Again, as is the case with length, the experimental results are dependent upon the type of task that has been performed by subjects. Letter frequency does appear to influence speeded tasks such as letter matching, naming and classification tasks (Balota, 1994). However, it does not appear to influence accuracy in perceptual identification tasks (Balota, 1994).

Letter processing is also intimately linked to the quest for the functional unit of word recognition. This old and still unresolved question, goes back at least a hundred years (Cattell ,1886). Cattell and others after him have shown that very briefly displayed letters are better recognised when they are presented in the context of a word or pronounceable nonword than when they are presented in isolation. This effect, as we have mentioned earlier in this chapter, became known as the *word superiority effect*. A disputed initial interpretation of this effect was that the whole word is the functional unit of recognition and not the letters composing it. Nowadays however, most models of word recognition favour the letter as the unit of word recognition. In this same context, letters are also investigated with regards to how abstract are their encoded representation. Is there any role to be played by the physical appearance of letters in the recognition process? How much is the reading process affected by differences in letter shapes (e.g., A and a)? This is the central topic of the debate between the holist and analytical researchers. Many of the issues only briefly mentioned in this section are dealt with at length in later chapters.

2.3.2 - Onsets and rimes

A common assumption is that spoken words are strings of syllables which in turn are strings of phonemes. This assumption leads to the view that monosyllabic printed words are parsed into units that correspond to one or more phonemes. However, a number of works now suggest that spoken syllables have a hierarchical rather than linear internal structure (Fudge, 1969; Treiman, 1989). This subsyllabic division comprises the onset and rime of a syllable. The onset of a syllable can be identified as the initial consonant or consonant cluster in a word. For example /c/ is the onset in *cat* , /cr/ is the onset in *crack*. The rime for a word involves the vowel following the onset and any subsequent consonants. For example, in *CAT* and *CRACK*, *AT* and *ACK* respectively would be the rimes. These units in turn are composed of phonemes, as is illustrated in Fig. (2.7).

The evidence for the view that syllables have a hierarchical rather than a linear internal structure had been adduced by both linguists and psycholinguists. The evidence includes constraints on the distribution of phonemes in syllables (Selkirk, 1980), errors in the production of speech (MacKay, 1972), and people's ability to learn word games that break syllables at various points (Treisman & Chafetz, 1987).

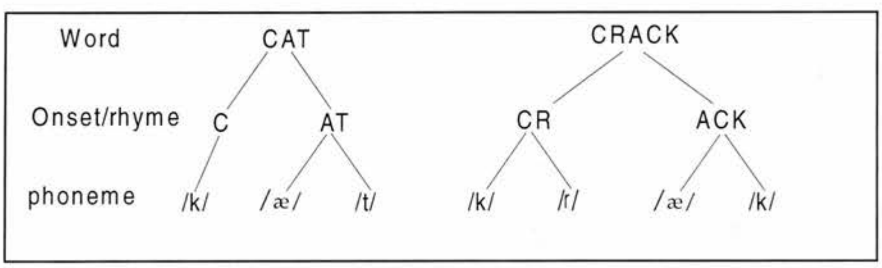

Fig. (2.7) - The hierarchical structure of words.

An important question for theories of word recognition concerns the nature of orthographic units in printed words and the extent to which these units mirror the phonological units of spoken words (Treisman & Chafetz, 1987). This was investigated in an interesting series of experiments where strings like FL OST ANK TR were presented to subjects who were then asked to determine whether two of the strings could be combined to form a real word. In the examples above, one can see that FL and ANK can be combined to produce FLANK, with FL corresponding to the onset and ANK corresponding to the rime of the word. Consider now the performance on conditions where the strings again correspond to words but they are not broken at the right places to form onsets and rimes. For example, a subject might have been presented with FLA ST NK TR. For these items, the correct answer is again FLANK, but now FLA and NK do not correspond to onsets and rimes. Treisman and Chafetz experiments indicated that anagram solutions yielded faster response time when the breaks corresponded to onset-rimes divisions compared to when the breaks did not. A similar pattern was found in an experiment where adults made faster lexical decisions in words presented with slashes inserted between orthographic onset and rime units (e.g. SL//IP) than in words with slashes inserted within orthographic units (e.g. SLI//P). In a later work, Treisman and colleagues went further claiming that letter groups that correspond to the rimes of spoken syllables, or units that include the vowel and any following consonants, play an important role in adults and children's pronunciations of printed words.

Support for the *rime* as having a functional role in word recognition is also to be found in experiments using the concept of word-body (Patterson & Morton,1985; Kay & Bishop,1987) which is analogous to the rime. The word body is defined as the grouping of letters in a monosyllabic word that consists of the terminal consonants (or coda) plus the vowel. For example, in the same way as the rime, the body of the word PLANT is ANT or the body of CAT is AT.

In summary, a hierarchical view of the syllable structure has replaced the old linear view among linguists. This notion has been extended from the realm of phonology to the realm of orthography. Finally, the experiments above are part of a considerable number of studies that indicate that orthographic rimes (word-bodies) may function as units of print.

We now move to the level of the word as a whole and describe briefly some variables that are important to the work that has been carried out in this thesis.

2.4 - Visual word recognition variables

2.4.1 - Frequency

Frequency is a ubiquitous variable of word recognition and the relevance of its effects is well acknowledged in the literature (Howes & Solomon, 1951; Foster & Chambers, 1973; Whaley, 1978; Balota & Chumbley, 1984; Inhoff & Rayner, 1986). Its influence has been shown in a variety of experimental paradigms such as naming, lexical decision, tachistoscopic reports, "same-different" matching tasks and semantic categorisation. Common high-frequency items, such as *table* are recognised more rapidly and/or accurately than less common low-frequency items such as *blare*.

The size of the effect has been shown to be dependent upon the task performed. Nevertheless, it is found to be robust and often large in many experimental situations. In the lexical decision task (LDT) paradigm, subjects are required to decide if a string of letters is a word or a nonword. The LDT paradigm exhibits one of largest frequency effects when compared to other paradigms. Experimental investigations using the naming paradigm also report frequent words to be named faster than rarer ones (Frederiksen & Kroll, 1976; Foster & Chambers, 1973). Generally however, the effect is reported to be smaller than that found in the lexical decision task (LDT) (Andrews, 1989).

The frequency effect is known to be pervasive and difficult to disentangle from other factors. Therefore, in any experiment it is essential to control the frequency of the items and also ensure that spurious factors, such as length, are matched when designing the experiments. Taking word length as an example, less frequent words tend to be longer than more frequent ones. So, when using these two types of words in frequency effects experiments, one must ensure that the items are matched for word length. Otherwise, the fact that less frequent words take longer to recognise than more frequent ones could also be attributed to the fact that less frequent words are longer than their counterparts.

So far the term word frequency has been employed assuming an intuitive understanding of its meaning. It is unlikely that the term would be misunderstood but it can be defined simply as the number of times a word is encountered by the users of a language. Its measurement is objective in the sense that the frequency is generally calculated from studies of the occurrence of a given word in a large body of text or transcribed speech. The most commonly cited frequency databases record the total number of occurrences of each word in corpora containing a few or a few tens of millions of word occurrences.

The vast majority of models of visual word recognition have been developed with the provision that the recognition process is sensitive to frequency (Monsell et. al., 1991). However, the locus of the frequency effect in word recognition has not been established yet. Are frequency effects inherent in the way words are stored or do they merely affect the way in which subjects respond in experimental tasks? Several lines of evidence have shown that processes subsequent to lexical identification, subsequent even to access to meaning or pronunciation, could be a good candidate for the locus of frequency effects. It has been found, for example, that frequency effects were stronger for the same material in a lexical decision task than in a semantic categorisation task (Balota & Chumbley, 1984). Since lexical decision and semantic categorisation both require lexical access, most of the frequency effect on the LDT must be attributed to the decision stage (after lexical access), specific to the lexical decision task. McCann & Besner (1987), after showing the absence of a frequency effect for pseudohomophones, together with evidence that they do activate lexical representations, concluded that the mere activation of the lexicon can not be the locus of the frequency effect for ordinary naming. The importance of locating the frequency effects resides in the fact this would allow a more precise determination of how the properties of a word affect its recognition. In addition, this would also allow a better evaluation of the strengths and weaknesses of the many models of word recognition that are in use at present.

Different interpretations of the frequency effect have appeared recently. Morrison and Ellis (1992), for example, argue that part of the effect attributed to frequency can be equated with the age of acquisition of words. High-frequency words tend to be learned earlier in life than low-frequency words, so that sets of words selected as being high- or low-frequency of occurrence tend also to be sets of words which are early- or late-acquired, respectively. They reported six experiments which contrasted the effects of frequency and age of acquisition on written word recognition. Age of acquisition affected word-naming speed when frequency was controlled but there was no effect of frequency when age of acquisition was controlled.

The variable of familiarity, has been used by some researchers as a substitute for frequency counts. They criticise frequency effects on the grounds that experiential familiarity counting takes into consideration a broader spectrum of the use of a word than frequency counting, as the only source for the latter in the majority of cases is texts.

2.4.2 - Experiential familiarity

Gernsbacher (1984) argued that familiarity is a more fundamental processing variable than frequency. Familiarity can be thought of as a measure of personal frequency. In a series of experiments she demonstrated that word familiarity accounted for the inconsistent interactions between word frequency and a number of other lexical variables (bigram frequency, concreteness and polysemy), particularly for low-frequency words. She suggested that experiential familiarity, as determined by subjective ratings may be a more comprehensive measure than printed word frequency, in that it may take into account all encounters with a given lexical item. For example, familiarity measures may reflect exposure to words from language production, as well as experience with auditory, visual and written forms. The interest in familiarity ratings was borne out of the observation that there is a great variation in the experiential familiarity of the low frequency words. Some words with recorded low frequency (such as *discard*) are rated more familiar than others of similar frequency (such as *groats)*. The disadvantage of the experiential familiarity variable is that it is intrinsically subjective (Taft, 1991). This is its major weakness compared to word frequency, where an objective measurement can be made. Taft, for example, argues that one of the problems in adopting subjective familiarity is the fact that raters could be basing their feeling of familiarity on a sense of how long it takes to recognise a word. As an example, he cites the case that if a rater is able to access a concrete word more quickly than an abstract word matched on objective frequency, the former might be rated as more familiar than the latter. This obviously, would result in a serious confusion of the concreteness and familiarity variable.

2.4.3 - Neighbourhood effects

The concept of orthographic neighbourhood has received considerable attention in word recognition and lexical access research. The 50 000 words estimated to be in the vocabulary of the average reader of English (Monsell et., 1989) are constructed from a limited number of 26 letters which are themselves made up from a small set of features (Gibson, 1969). Necessarily then there must be a considerable overlap in spelling patterns across words. Perhaps the most important issue in models of lexical access, concerns how the access mechanism selects the correct lexical entry to form a set of plausible candidates. The most widely used operational definition of neighbourhood density, states that it is equal to the number of words that could be generated by changing only a single letter in any of the positions within a word (Coltheart et. al., 1977). Thus, the words *trick*, *crack* and *trace* are all neighbours of the word *track*

Next we discuss the two factors that need to be accommodated when discussing neighbourhood effects. First, one needs to consider the influence of neighbourhood size. Some words are embedded in relatively large neighbourhoods, whereas others are embedded in relatively small neighbourhoods. Specifically, in both pronunciation and lexical decision performance, low-frequency words from large neighbourhoods produce faster latencies than low-frequency words from small neighbourhoods, whereas there is little or no influence of neighbourhood size for high frequency words

(Andrews, 1992). Second, one might also expect the frequency of the neighbours to play a role in word recognition tasks. There have been some arguments that neighbourhood frequency is even more important than neighbourhood size (Grainger, 1990; Grainger, et. al., 1989; Grainger & Segui, 1990). Performance was predicted to be worse whenever the target word has at least one orthographically similar word that is higher in frequency. Because the likelihood of a higher frequency neighbour should increase with the size of the neighbourhood, results of Grainger seem to be inconsistent with those obtained by Andrews (1992). However, this inconsistency can be resolved, according to Andrews (1992) due to two facts. First, that the difference between the size of the small and large neighbourhoods was smaller in Grainger's studies than in Andrews' studies. Also, Grainger has only found this inhibitory effect in lexical decision, and hence it is possible that this pattern may reflect decision processes that have been tied to this task.

It would seem that both frequency of neighbours and size of neighbourhoods should play a role in word recognition tasks. Luce and Pisoni (1989), for example, when using auditory priming task found the number of competitors, which they call the neighbourhood density, influences the decision. The probability of identifying a stimulus word, according to them, is equal to the probability of identifying of the stimulus word divided by the neighbours. They also argue for differences across experimental tasks such as lexical decision, pronunciation and threshold identification tasks. Marslen-Wilson (1990) examined the effect of frequency of competitors upon recognising words and he found that the time it takes to recognise a word such as "speech" does not just depend on the relative uniqueness points of competitors (such as "speed" and "specious") in the cohort, but also on the frequency of those words. Hence, people are faster to identify a high frequency word which only has low frequency neighbours[1] than vice versa. The rise in activation of high frequency word is much greater than for a low frequency one.

Finally, neighbourhood size effects have been considerably effective in discriminating between the search models and the activation models of word

[1] Here the concepts of neighbourhood and cohort were collapsed into one definition meaning "competitor set".

recognition. Neighbourhood size effects would appear to produce particular difficulties for serial search models. Specifically, the more items that need to be searched, the slower the response latency should be. This is precisely the opposite of the pattern reported by Andrews (1989), who finds that larger neighbourhoods produce faster response latencies, and only for low frequency words. However, Seidenberg and McClelland (1989) have demonstrated that their connectionist model of reading aloud, can nicely accommodate Andrews' effects of neighbourhood size.

2.5 – Stimuli

So far this chapter has discussed the relevant models and variables involved in visual word recognition. In the next section some words will be said about the stimuli used in the experiments carried out in this work. As both nonwords and brand names were used, some of the literature and issues involved will be discussed. All the experiments run as part of this thesis had either nonwords or brand names as their stimuli material. High frequency common words were also used in experiment 4 and 5. The aim of this section is to give a brief description on both type of stimuli. Below nonwords are discussed first and next brand names are introduced.

2.5.1 – Nonwords

Dealing with novel strings of letters, that have never been encountered by the reader before, is of fundamental importance in reading and pronunciation. Texts can be a mixture of familiar and unfamiliar words even for the skilled reader. The ability to make sense of these nonsense strings and to assign pronunciation to a new pattern of squiggles, never ceases to be an important skill in reading.

It would be extremely difficult, if not impossible, to find a set of adult skilled readers all having the same learning experience to act as subjects in psycholinguistic experiments. This would be necessary to ensure that the novel words would be novel for all subjects. Nonwords were introduced in psycholinguistics as a simple solution to this problem. These are strings of letters arranged in such a way that they do not correspond to any known word in the language. Although sometimes recognised as a fairly artificial type of stimuli, their invaluable contribution to the modelling of visual word recognition has always been unanimously acknowledged.

The *lexical status effect* is well known and consists in the fact that words are responded to faster in experimental situations than nonwords (Rubenstein et.al., 1970). Subjects, for example, take longer on average to respond to the string "lece" than to the string "lace". One feature of nonwords is that they can be created to have the same orthographic structure of their real counterparts that are normally found in the mental lexicon. Nonwords with this characteristic are normally termed *pseudowords*. One method of creating pseudowords is for example, to take a real word and change only one of its letters so as to make it a nonword. This is exactly the case of the example given above, where the letter *a* of the word *lace* was replaced by the letter *e*, creating the nonword *lece*. However, nonwords can also be created by randomly putting letters together, so that the final arrangement does not obey the structure of the language. Obviously, this will produce many *orthographically illegal* nonwords, for example, the string *lcee*. The latter group is also called *non-pronounceable* nonwords. Response times are longer for non-pronounceable nonwords compared to pseudowords, which became known as the *nonword legality effect*.

Another relevant effect involving lexicality is the *word similarity effect*. The closer a pseudoword resembles a real word, the more difficult it is to classify it as not being a word. This type of effect is observed for example, when a nonword is created by transposing letters from a real word, as it is the case for example with the pseudoword "trian" (train) when compared to "truan", which is a nonword generated in a different manner. Finally, there is the *pseudohomophone effect*. Pseudohomophones are nonwords which sound like words when pronounced (such as *brane*, which sounds like *brain* when spoken). The behaviour of the pseudohomophone *brane* can be compared with the very similar nonword *brape*, which does not sound like a word when it is spoken. It has been shown that pseudohomophones are more confusable with words than are other types of nonwords (Rubenstein, Lewis & Rubenstein, 1971). Participants are faster to name them, but slower to reject them as nonwords than control nonwords. The pseudohomophone effect does appear to be real, but it is found only in restricted conditions (Laxon, Masterson, Pool & Keating, 1992).

In short, we have seen above that there are many different ways of creating nonwords. This leads to the grouping of nonwords in many different categories, like pseudowords, non-pronounceable words, pseudohomophones, etc. Depending on which categories of nonwords are used in experiments, different effects are obtained.

2.5.2 - Brand-names

Brands are taken very seriously by those working in marketing strategies and are defined as follows:

> "A name, term, sign, symbol or design, or a combination of them which is intended to identify the goods or services of one seller or a group of sellers and to differentiate them from those competitors."[2]

Brand names are the part of the *brand* that can be vocalised. It is in many cases the usual means of asking for the product one is interested in. The search for appropriate brand names is a multimillion dollar business. As a consequence, a lot of research is carried out that investigates the consumer

[2] Marketing Definitions: A Glossary of Marketing Terms (Chicago: American Marketing Association, 1960).

behaviour to brands. Part of that investigation is directed towards research on brand names exclusively. However, as we show in Chapter 4, it is only recently that cognitive psychologists have become interested in the study of the category of proper names, of which brand names are a subcategory. This interest has partly been promoted by the research carried out on face recognition models. Insofar as we are aware, work with brand names from a cognitive perspective is almost non existent. An exception to be mentioned is the work reported by Rubin et. al. (1991), where they were interested in prototypical behaviour connected to the names of certain type of products. A more detailed account of this work will be given in Chapter 4.

There are a number of reasons why brands are an interesting type of stimulus to investigate. They possess a rich array of inherent perceptual features, such as colour, size, case, fonts etc. In the present context we focus on the influence of case familiarity in the recognition of brand names. The impact on our cognitive abilities of these perceptual features in the context of advertising has not yet been investigated. One factor that makes the investigation of brand-names attractive is the constancy of their presence in our daily lives, in both modalities visually and auditory, due to aggressive advertising. Furthermore, brand-names suit the ideal type of psycholinguistics stimuli. They can, for instance, be tightly controlled in terms of orthography, phonology and semantics. Academic books on advertising, state that in the hunt for a successful brand-name, every aspect of the string to be used in a campaign, must be considered, such as their number of letters, syllables and semantic associations (Frey & Halterman, 1970).

In marketing brand-names are classified in seven categories. Below they are listed together with examples:

1- names of companies - General Electric refrigerators or Heinz varieties.
2- Combination of letters in a manufacturer's or company name - ARMCO (American Rolling Mills), SOHIO (Standard oil of Ohio).
3- Coined names: invented by their owners - KODAK. Names representing the misspelling or simplified spelling of one or more words. *Takhoma, Frigidaire, Durez, Deepfreeze, Formica.*

4- Arbitrary combinations of letters or numerals - B.V.D. 7-20-4.

5- Names of places, people and characters - *Paris, Lincoln, Cinderella.*

6- Names suggesting quality or good performance - *Royal, Perfection, Ideal.*

7- Easily remembered symbols - *Arrow, Diamond*

As it can be seen from above, there is a large degree of freedom in the manipulation of language for the creation of brand names. This provides a unique opportunity for studying aspects such as the memory for strings of letters belonging to different contexts.

Another area open to investigation is that of neuropsychology of proper names. There are many reported cases in the neuropsychological literature of patients who have for example, after a stroke, lost the ability to recognise proper names, despite having kept intact their ability to recognise all other category of words. This will be discussed in more detail later on in Chapter 4. However, up to now, neuropsychologists seem to have paid little attention to the investigation of the sub-category of brand names.

Old books on advertising, published thirty years ago, but still consulted, contained statements such as:

> "ALL CAPITAL LETTERS ARE HARDER TO READ than lower case letters such as those used here. Condensed letters are harder to read than set in regular width"[3]

Experimental findings confirm the statement above by showing an small advantage for reading lower-case over upper-case print (Woodworth, 1938; Smith, 1969; Fisher, 1975). The obvious thought that the appearance of a string can influence its reading misled scientists in thinking at the time that this was one of the more treatable areas in reading. Evidence can be seen, for example, from this passage extracted from Brooks (1977):

[3] Kleppner, O. (1966). Advertising Procedure. Prentice-Hall, New York.

> "Many of the important aspects of the reading process are frustratingly difficult to unravel. But in this sea of intangibles, there is at least one variable that we can easily manipulate: the visual form of words."

In reality the statement was too optimistic. Twenty years later, and there is still much controversy about the role played by the visual characteristics of words in reading. In Chapter 6 I address the psycholinguistics issue relevant to brand names and present the results of a series of experiments which investigate this issues.

2.6 - Summary

In this chapter, the concept of the mental lexicon and its importance for the theory of visual word recognition were introduced. Some of the most influential models of visual word recognition were quickly discussed and then we saw how theories of visual perception (based on pattern recognition processes) help to explain the initial stages of visual word recognition. The fact that some of the constituent parts of the words (letters and onset/rimes), can be regarded as plausible candidates to fulfil the role of functional units of word recognition was also considered, followed by a discussion of some of the variables that influence visual recognition at the level of the whole word. Finally, in the section above, nonwords and brand-names, were introduced, as they were used as stimuli in the experiments that will be presented later in this thesis.

Chapter 3

General Methodology

Philosophy of science has demystified the naïve belief that natural science, and for that matter all sciences, can be a hundred percent objective in their observations and theories. Despite this, objectivity has remained one of the prime goals of scientists. Objectivity and methodology are intrinsically linked, with the latter determining the degree of the former. One way of understanding the role of research methods is to see them as a means of minimising the probability that observers will distort facts.

In the present chapter, I describe the methodological approach taken in this thesis. I start by discussing very briefly the method of presenting isolated words to subjects in reading studies. Next, I describe the type of stimuli used in the experiments, as well as the experimental tasks that were involved. Finally, I give an account of how data measurements were taken.

3.1 - The presentation of isolated words

Researchers are well aware of the fact that the presentation of isolated words on a computer screen is a fairly artificial method of experimentation, since normal reading happens most of the time in the context of sentences that are generally part of a text. However, the method is well established and has proved to be very fruitful in a good number of the experimental situations. Its inherent tractability outweighs the disadvantage mentioned above. The exclusive adoption of the sentence or even the whole text approach would bring so many new variables into play that it would render objective studies almost impossible.

Henderson (1982) reminds us that the study of words in isolation can be justified as well, by considering for example, the fact that the teaching of reading starts with isolated words. The ability to arrive at the name of isolated words turns out to be a surprisingly powerful diagnostic test of backwardness in learning to read, as well as in various types of acquired dyslexia consequent on brain damage. Also, decisions at the word level play a fundamental role in almost all theoretical models of reading and of speech perception.

Finally, it is generally held by the majority of cognitive psychologists that "words are processed pretty much the same way in isolation as in text" (Pollatsek & Rayner, 1980). Note however, that this is not the same as saying that findings with isolated words can be assumed to be perfect indicators of how word recognition operates in reading text.

All the experiments described in this thesis involve the recognition of strings of letters that are presented in isolation on a computer screen. Experiments 1 and 2 (Chapter 4) are naming tasks where nonwords are used as the stimuli material. Experiment 3 (Chapter 5) used a same-different matching task where pairs of nonwords were simultaneously presented on the screen. In Experiment 4(Chapter 6), a lexical decision task was again used, having common English words, brand names and nonwords as stimuli material. In Experiment 5 (Chapter 6) only brand names were used in a naming task. Finally, in Experiment 6 (Chapter 7) nonwords were phonetically transcribed by a group of phoneticians.

3.2. - The algorithm for creating English nonwords

In section 2.5.1 of Chapter 2 we discuss the importance of the role played by nonwords in psycholinguistic studies. In section 2.4.1 of that same chapter we describe the relevance of the frequency variable. There we call attention to the fact that strictly speaking frequency counts can not be attached to nonwords, since by definition they are not supposed to exist as such in the language. The traditional method of creating nonwords is to have them derived from real words. Nonword frequency is assumed to be somewhat related to the frequency of the real word from which it has been derived. Below, we describe an alternative method for creating nonwords. This method has been inspired by the quasi-human performance on nonword reading achieved by Plaut et. al. (1996) neural net. The authors have argued that the neural net's success was partlty due to the fact that they had encoded information in the net by using the concepts of onset, nucleus and coda. The reason this method has been chosen in place of the more traditional one is that controlling the frequency of the parts that together contribute to form a nonword is crucial to some of the investigations we have carried out in this thesis. Also, some psychological reality is attached to it due to the use onset, nucleus and coda that are known to affect reading.

The algorithm for building nonwords to be presented here, was based on the principle that monosyllabic words are composed of two units, i.e., the onset and the rhyme (word-body). This issue has been reviewed in the previous chapter at section 2.3.2. Next, we present the steps used in developing the algorithm.

Step 1:

Initially, a file containing all English monosyllabic words and their frequencies of occurrence was created by extracting this information from the English language lexical database CELEX[1] (Baayen et. al., 1995). The frequencies had been calculated from a 17.9 million word corpus of English words (COBUILD corpus).

[1] CELEX lexical database of English (version 2.5) developed by a joint enterprise of the University of Nijmegen, the Institute for Dutch lexicology in Leiden, the Max Planck Institute for Psycholinguistics in Nijmegen, Institute for Perception Research in Eindhoven

Step 2:

The next step was to extract from this file all possible combinations of letters composing the onset, nucleus and coda of monosyllables. Next, three separate files were created, one containing only the onsets and their frequencies (onset-file), another for the nucleus and their frequencies (nucleus-file) and finally one for the codas and their frequencies (coda-file). The information in the three files was extracted with the help of small grep/AWK programs (in the UNIX environment) that searched for the required data from the monosyllable file. An example of one these programs can be seen below.

```
grep '\< br' freqmono.txt / awk '{ c = c + $2} END { print "br", c}'
```

In this program, the onset "br" is being searched and freqmono.txt is the file containing the monosyllables and their frequencies. For every line of freqmono.txt, grep tests if the present word begins with "br". If so, AWK is invoked and the frequency of the present word is added to the variable "c". After all lines (i.e. all monosyllables) of freqmono.txt have been tested, the variable "c" will contain the number of occurrences of the onset "br". Thus, the final instruction in the program prints the onset "br" and its corresponding frequency. To test other onsets, they are simply substituted into the program above, in place of "br". Similar programs were used to create the nucleus and coda files.

Step 3:

The three files were then sorted into descending frequency order and the onsets, nuclei and codas where classified accordingly as having high or low frequencies. The onset file was composed by a list of 78 different onsets. Most lexical databases contain a column that gives the frequency of a word in

1 million counts. It is generally agreed that it makes greater sense to say that the frequency of a word is one in a million than it does to say that it is 22 words out of 16 600 000. However, it is also the case that some detail is lost in this scaling-down process. For example, information on very low-frequency words is lost as the resulting frequency will be truncated to zero. For that reason, we opted here for having the frequencies of our onsets, nuclei and codas sorted out on the basis of the totality of the database, i.e., 17 900 000 words.

The onset frequencies ranged from 184960 to 1. The cut off point between the high and low-frequency was arbitrarily chosen in the following way: those onsets which had less then 5000 appearances in the database were regarded as low-frequency. This corresponded to 27 onsets (34.6%) of the total. The high-frequency onsets numbered 51 and corresponded to (65.4%) of the total.

The nucleus file was composed of 21 combinations of pairs of vowels[2]. Their frequency ranged from 455035 to 2. The classification of vowel pairs into low or high frequency was not taken into account in the algorithm.

The total number of codas found in the CELEX database was larger than the number of onsets. A total of 185 different codas were found, their frequency ranging from 1003914 to 1. The cut off point used between high and low frequency for codas was set at codas of less than 100 appearances in the database. This corresponded to a total of 43 codas (25.2%). The rest 142 codas (76.8%) were classified as high-frequency. This information was then used in building the nonwords.

Below is an example of how to build tightly controlled nonwords regarding the frequency of their onset and codas. For example, for the high-frequency onset - high frequency coda (high-high for short) nonword HOUT, we have:

[2] Pairs of vowels were used to compose the nucleus because they help to increase the number of possible pronunciations per nonword.

subsyllables	**HOUT**	frequency
onset	h	789206
nucleus	ou	455035
coda	t	1003914

Table (3.1) - A high-frequency nonword is created by means of the algorithm above.

For a low frequency onset - low frequency coda nonword (low-low for short) such as SVAANTZ, the frequencies are as follows:

subsyllables	**SVAANTZ**	frequency
onset	sv	4
nucleus	aa	81
coda	ntz	10

Table (3.2) - A low-frequency nonword is created by means of the algorithm above.

As seen in Chapter 2, there are a number of ways of creating nonwords. The procedure used here for creating nonwords is different from those we have seen in Chapter 2. The advantage of the present method is that we can accurately control the frequency of the parts forming a nonword. This can be taken as a measure of the "nonword frequency" itself. This is an improvement on previous methods of creating nonwords, where no such degree of control is possible. In the next section the frequency variable is considered in more detail.

3.3 - Stimulus specifications

3.3.1 - Frequency

Frequency is generally defined, as in Chapter 2, as a value that can be assigned to a word regarding its number of occurrences in a language. It is a very important variable in psycholinguistic experiments known to influence many experimental results. The frequency with which a word occurs in printed English is a consistent predictor of response time in tasks such as lexical decision and naming (Balota & Chumbley, 1973; Frederiksen & Kroll, 1976). However, according to the above definition, it is meaningless to talk about the frequency of nonwords. These are strings of letters that are not found in the lexicon, but are created by cognitive psychologists as a means of experimenting with novel stimuli. Researchers get around the fact that frequency counts can not be attributed to nonwords per se, by adopting as a method for creating them, the device of taking a high- or low-frequency word and then changing one of its letters so as to have a nonword as the result. Take the high-frequency word *home* for example; by changing the letter *o* to the letter *e*, the end result is the nonword *heme*. In the case of the low-frequency word *pork* by changing the letter *o* to the letter *u*, the end result is the nonword *purk*. Now, the two nonwords can be meaningfully contrasted, for example in terms of the time taken to react to them. In the seminal paper published by Glushko (1979), he uses this method for controlling the frequency of his nonwords. Frequency effects regarding both words and nonwords are a prominent feature of almost all the word recognition models in the literature.

The technique for nonword creation described in our algorithm seems to be a better and more objective way of controlling the frequency variable than the method described above. With our algorithm, because the subsyllables are manipulated according to their frequency of occurrence in a large database, there is no danger of creating strange letter combinations, when the intended nonword was supposed, for example, to have a "high frequency". Moreover, our algorithm allows the creation of nonwords that exhibit four different types of frequency combinations. They can be classified as high-high, low-low and high-low, low-high frequency nonwords. The majority of the nonwords used in our experiments were of the first type(high-high) nonwords. Only in Experiment 1(Chapter 4) the low-low type of nonword was used. The reason for not using low-low nonwords beyond that of Experiment 1 is that most of them would not be orthotatically similar to common English words and this factor could interfere with the frequency effects we were investigating.

The common English words used in Experiment 4 (Chapter 6) have also been controlled for frequency. Only high frequency words, with more than 50 occurrences per million entries in the CELEX database were used (Riddoch, Humphreys, Cleton, & Fery, 1990).

3.3.2 - Experiential familiarity

Apart from very common people's names, frequency counts for proper names are not available or very hard to obtain. Counting the occurrences of personal names in telephone directory books has been used as an estimate for the frequency of names of people (Valentine et. al., 1991). In the case of brand names, frequency counts are even more difficult to obtain, if not impossible currently. As a solution to this problem we chose instead to ask a group of people to rate in a scale from 1 to 7 how much they thought a determined brand name to be familiar to them. Details on how this rating has been obtained can be found in section 6.4.1.2 of Chapter 6.

3.3.3 - Weirdness

The effect of weirdness was explored for different type of tasks in Experiment 1, 2 and 3. The nonwords created are a combination of onset +

nucleus + coda , where each of these subsyllabic parts belong to the English lexicon. A nonword is defined as *weird,* when the combination of its nucleus + coda can not be found in the lexicon. The *nonweird* nonwords are those whose nucleus + coda are found in the lexicon. These classifications are founded in the literature of syllables discussed in the previous chapter. An interesting feature of the weirdness variable is that weird nonwords tend to be neighbourless and by contrast the nonweird nonwords can sometimes possess a number of orthographic neighbours. This distinction is important because as mentioned in section 2.4.3 of the previous chapter, the size of the neighbourhood of a string affects the way it is read.

3.4 - Experimental tasks

One of the crucial differences between experimental studies carried out in psycholinguistic research and those carried out by other branches of science, the so-called "hard-core" sciences, as physics, for example, is the nature of the investigation involved. The results from psychological experiments are more open to multiple interpretations and varying structure/process tradeoffs than those belonging to the branches of science referred above.

The tasks used in investigating word recognition have come under much scrutiny for their ability to reveal the workings of the organisation of the lexicon. Tasks vary in suitability depending on the questions asked. We cannot take for granted that all subjects perform a particular task in a given way that maps directly onto an assumed functional architecture. For example, when performing lexical decisions, subjects may check whether the letter string is meaningful or whether it is visually familiar (Besner & McCannan, 1987). Although a task, may, in theory, tap a particular processing stage, in reality performance may be based on a number of processing stages whose precise contribution may differ on a trial-by-trial basis. Also, each task is differently affected by a host of different variables that play a role in psycholinguistic research. Below, the tasks that were used in the experiments carried out in this thesis are presented.

3.4.1 - Lexical decision

This task is perhaps the most frequently adopted task in the study of lexical access. It requires the subjects to decide if the stimulus being presented to them is a word or not. The most commonly used version of the lexical decision task (LDT) and also the one used in the present work, is visual, where words are presented on a computer screen. There are however others, for example, auditory versions of the lexical decision task.

The LDT has been in use for almost thirty years (Rubenstein et. al., 1970, 1971) and is considered to be an appropriate method for studying word recognition, because in order to perform the task, subjects need only to access their mental lexicons. No further processing of the letter strings, such as identifying meaning, is required. Speed and accuracy of response are used to measure the difficulty of lexical processing. However, the absolute response time measured is in general, not very informative. Differences in response times, caused by the manipulation of experimental variables give a direct measurement of the influence of such variables in the experiment.

However, ever since decision studies began to be reported, the lexical decision task has been known to be a highly frequency sensitive measure (Rubenstein et. al., 1970) and known to show stronger frequency effects than the naming task (Balota & Chambers, 1984). There is also quite a lot of evidence that lexical-decision time is influenced by semantic properties of words. James (1975) obtained a faster average lexical decision time for concrete than for abstract low-frequency words matched for frequency and other properties. Jastrzembski (1981), and Millis and Button (1989) have demonstrated substantial effects on LDT of the number of meanings possessed by a word, although this effect is not uncontroversial (see, Gernsbacher, 1984). In Whaley's (1978) multiple regression study of lexical decision, the inclusion of various "richness of meaning" variables (concreteness, meaningfulness, imagery) accounted for significant variance after the effects of word frequency and word length had been removed. Neighbourhood effects are also known to be strong in LDT being some times inhibitory and some time facilitatory(Voice, 1995).

The last issue concerning the LDT that we are going to present has some bearing on our motivation for using this task in the present work, i.e., the questioning in recent years of the assumption that the lexical decision task taps directly and transparently into lexical identification processes. Balota & Chumbley (1984) have proposed that, lexical decision involves a nonlexical process based on the evaluation of letter strings along a familiarity-meaningfulness (F-M) dimension. Fast response are made if the computed value falls below a low criterion or exceeds a high criterion. Further processing is needed if the F-M value falls between the two criteria. This opens the possibility that the decision processes involved in lexical decision might be the locus of various effects previously attributed to lexical identification. For example, effects of word (DE Groot, 1984) or sentence (Foster, 1981) context upon LDT have been attributed to the decision process being biased toward "yes" or "no" by the post-lexical access detection of "congruence" and "incongruence" between a word's sense and that of the context. Feustel, Shiffrin, and Salasoo (1983) suggested that the facilitatory effects of repetition on LDT were due to the decision process being biased by the "familiarity" of the letter string, detected on the basis of retrieval of an episodic memory trace of a prior encounter with the letter string. In relation to frequency effects Balota & Chumbley (1984) and Besner & McCannan (1987) attribute a major component of the frequency effect in the lexical decision task to a familiarity discrimination process distinct from lexical identification. Although it is universally accepted that lexical decision is a kind of familiarity judgement the claim that frequency effects in lexical decision have a "post-lexical" locus can not be regarded as conclusive.

The two major motivations for using the LDT task in our experiments are because of its prolific use and established reputation in the cognitive psychology literature, it allows reliable comparisons between previous experiments and the ones that were run in the thesis. Also, as the literature shows, it is the ideal type of task to investigate the contrasting effects of capitalisation that are related to the identification and figural familiarity processes in visual word recognition.

3.4.2 - Naming

In the naming task individual strings of letters are visually presented to the subjects, generally on a computer screen. They are then asked to name aloud the presented string. The time taken between the presentation and the beginning of articulation is then measured. It is generally accepted that naming tasks tap the subjects' automatic access to the lexicon better than the lexical decision task (Lorch, Balota & Stamm, 1986; Neely et. al., 1989). Like theories of the LDT, theories of naming are based on the notion of activation; the assumption underlying naming is that highly active concepts are more available for pronunciation, and thus positive targets are named more quickly (Potts et. al. 1988, Seidenberg et. al. 1982, 1984). Because naming does not involve decisions, the criticisms of the decision methods do not apply here. Presumably these techniques do not elicit strategies like the plausibility strategy and criterion shifts (Albrecht & O'Brien, 1991 ; Keenan et. al., 1990). Naming also has an advantage in terms of the criterion of naturalness; pronouncing a word is more natural to subjects than having to decide whether a target is actually a word or not (Foster, 1981). The incidence of errors is relatively small (< 3%; see Warren, 1977).

However, naming tasks also present problems. Critics have observed that naming procedures may be of limited use in detecting inferences for theoretical and methodological reasons. Problems may result from the fact that naming procedures are based on the articulatory system and its specified mechanisms. Articulations, including naming, tend to have short latencies and entail the possibility of floor effects. Unless target words across conditions are identical, researchers must be mindful of what Steinberg and his colleagues (1980) have called articulatory unpacking and execution. According to Sternberg, articulations are planned in working memory, the more complex the structure of the articulation, the longer it takes to initiate it. Appropriate control words must therefore be chosen to match the syllabic structure of experimental target words.

Another potential limit of the naming paradigm is that strings can be pronounced according to rules of grapheme-phoneme correspondence without retrieving meaning (e.g. Foster, 1981, Halle, 1990). Although most experimentalists believe that the naming task does routinely involve lexical access (Foster & Chambers, 1973), it is possible for many letter strings to be pronounced without recourse to lexical access. Regular words like MINT, for example, could be pronounced by means of simple rules that convert letters to sounds, rather than by means of lexical information. In this respect the lexical decision task might be a more appropriate tool for the investigation of lexical access. Furthermore, although the naming task seems closer to pure lexical access than the lexical decision task, it can not be seen as simply accessing the lexicon. Naming times also include the time it takes to access the sound of a word, its phonological code, after the word has been identified. Ultimately, this means that the naming task is sensitive to a different set of factors than those of the lexical decision task. Henderson (1982), for example, argues that word length affects naming but not lexical decision.

One of the crucial claimed differences between the processes involving the naming task and the lexical decision is that, whereas the identification of a string is a necessary step to the accomplishment of the former, it is not necessary to the accomplishment of the latter task (Besner & Johnston, 1989). This distinction is at the root of the debate about whether visual familiarity influences word recognition. This distinction has been the driving motivation behind our experiments with brand names where we have tested the same hypotheses using both paradigms: the lexical decision and the naming task .This issue will be discussed at length throughout the thesis. The naming paradigm was used in Chapter 4 (Experiments 1 and 2) and in Chapter 6 (Experiment 5) to investigate capitalisation influence on visual word recognition.

3.4.3 - Same-different matching task

In a graphemic comparison task, the subjects are instructed to classify pairs of letter arrays as being the "same" if they contain only matching letters in corresponding locations, and as "different" if one or more locations contain different letters. Studies of the relationship between the "same" and "different" reaction times and such factors as the size of the letter arrays and the number and locations containing different letters have cast valuable light on the nature of mental comparison processes. This task has also proved useful in the study of "word superiority effects". It is now well-established that graphemic comparisons are facilitated when the stimuli are words, regularly spelled non-words or familiar abbreviations.

Unlikely the naming and the lexical decision task, the same-different matching task is not as popular a paradigm and in consequence there is less that can be said about it. Most of the experiments that have used the paradigm are from the 70s and 80s and they investigated visual familiarity effects in word recognition. The motivation for its use Experiment 3 (Chapter 5), is two fold, first it offered the perfect opportunity for distracting subjects from the main variable being manipulated without arousing much suspicion. Second, and perhaps most important, is that subjects seem to be able to accomplish the task by making using only of figural familiarity analysis, thus facilitating the appearance of "shape" effects where appropriate. The same is not true for tasks, such as naming and categorisation, where the identification of the letters forming the string must necessarily be done.

3.5 - Equipment

The experiment generator used for designing and running the experiments in this thesis was PsyScope (Macintosh software) versions 1.02 and 1.1.beta 9 (Cohen, McWhinney, Cohen, Flatt, & Provost, 1993). A button box was attached to the Macintosh computer and operated as the interface between the subject and the computer. Programs were written by myself during this work to control experiment presentation and data gathering.

The button box illustrated in Fig. (3.1) below produces very accurate measurements of the time interval between the presentation of a stimulus on the computer screen and the pressing of one of its buttons.

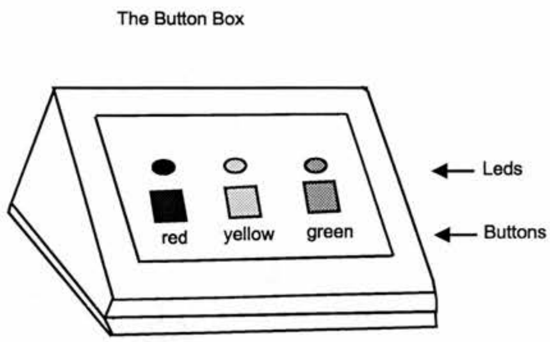

Fig.(3.1) - The PsyScope button box used for recording and timing responses made by subjects during psychological experiments.

It has its own built-in microprocessor, timing circuitry and computer interface and its accuracy is quoted by the manufacturers as being 1 ms. In addition, when the software initialises, the accuracy of the box is tested and synchronisation routines between the computer and the box are executed. The psycholinguistic programs are run only after the box succeeds in producing the desired accuracy of 1 ms.

Another useful feature of the button box is that a directional microphone can be coupled to it. In this case, the audio signal produced by the microphone can be used to trigger timing sequences or other actions in the software, in close analogy to the behaviour of a button in the box. Also, a threshold level can be set for the level of sound, so that noise can be discarded and prevented from triggering responses in the software. This combination of microphone and button box was used in the naming task described in chapter 6.

Apart from Experiment 1, all the onset latency measurements performed in the experiments described in this thesis were done using the button box . In Experiment 1 (Chapter 2) however, a different method was used to measure the onset latency of the nonwords. The experiment was designed so that before the appearance of each nonword on the screen a 350 ms (audio) beep was produced by the computer. The nonword was subsequently presented and the subject named it aloud, after a certain time interval. A directional microphone and a digital tape recorder were used to record the whole experimental session. The next step was to use the Entropic xwaves+ software (in the UNIX environment) so as to have the digitised speech of each beep + nonword displayed on the screen of a computer as a sound waveform. Onset latency was then measured from the onset of the beep sound to the onset of the pronounced nonword. Later the 350 ms of the beep duration was subtracted from the total reaction time obtained for each nonword. The method is illustrated by Fig. (3.2) below.

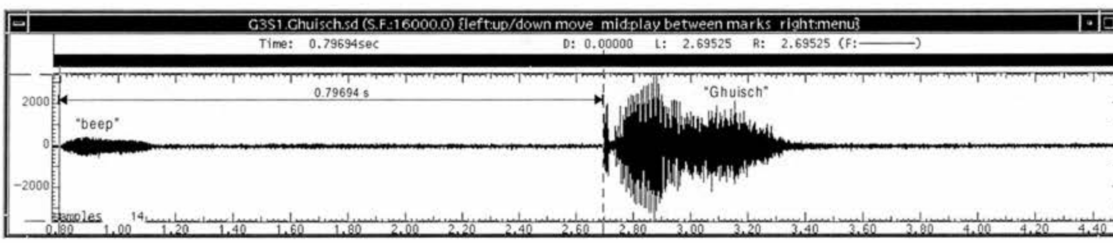

Fig. (3.2) - Waveform of the beep-sound and the nonword "Ghuisch" pronounced by subject G3S1. This figure shows how the reaction time was measured between the onset of the beep and the onset of the nonword.

3.6 - Measurements

The process of assigning numbers and units to particular features of objects or events is one of the foundations of the scientific method. Measurements allow researchers to accomplish three functions: description, comparison and prediction. The most fundamental reason to quantify observation is to allow the researcher to describe what occurred. Later examination of permanent records of experiments leads to new theories and ideas.

3.6.1 - Onset latencies

Onset latency or reaction time, as it is also known, is probably the most used type of measurement in psycholinguistic experiments. Part of its popularity derives from the quantifiable nature of the data it generates which in turn allows sophisticated parametric statistical tests to be applied. The straightforward manner with which reaction times can be gathered nowadays, due to computer controlled experiments has also contributed to their popularity.

Reaction times are simply a measurement of the time taken by a subject to react to an experimental stimulus. It is used to measure performance in a wide range of experimental tasks, as for example, in naming, lexical decision and categorisation tasks. The assumption behind its use is that the processes engaged during the tasks, each take a finite length of time, so that the duration of particular processes and their relations to other processes are revealed by the time subjects take to respond. Onset latency was the type of measurement used throughout this thesis, with the only exception of Experiment 2 (Chapter 4), where instead, we chose to count the number of different pronunciations produced by subjects.

3.6.2 - Number of different pronunciations

Ideally one would record all and only the variations between sounds of spoken words that cause a difference in meaning. The materials of experiment 1 and 2 were transcribed in broad phonetic transcriptions using the International Phonetic Alphabet (IPA). Any difference at the segmental level was recorded and treated as a prominent variant. Various pronunciations were produced in Experiments 1 and 2, for the string *baif*, for example. They were transcribed as follows: /beIf/ /bɛf/ and /baIf/. No suprasegmental differences, such as stress, length, tone and intonation were used.

The number of different pronunciations produced by the subjects, for each nonword was measured in Experiments 1 and 2. It will be seen later that this is sensitive to the variable being manipulated in those experiments.

Chapter 4

The implications of initial capitalisation for models of reading aloud

One of the aspects of reading that has been intensively researched in cognitive psychology is that of pronunciation. Information processing models and also computational models of reading aloud have been put forward that try to capture all its nuances. However, the ultimate model of reading aloud has proven to be quite elusive. This chapter is about one of the possible subtleties involved in the reading process: the role played by initial capitalisation as a clue in the reading of proper names. I start by briefly reviewing some of the work that has been carried out on the subject of proper names. Next, I describe in a broad manner the general status of pronunciation research in the field of cognitive psychology. I also describe two naming experiments undertaken during this work and report their results. Finally, I propose an explanation for the experimental results in terms of the *set size plausible phonologies* mechanism, that was originally put forward to explain why proper names are harder to recall than common words. I further discuss the implications of these issues in relation to connectionist models of pronunciation and suggest some adjustments to their architectures.

4.1 - Proper Names

The expressions *proper name* and *proper noun* are generally used interchangeably (c.f. Quirk et.al.,1985). Defining *proper names* is not a straightforward matter. There is some debate concerning the types of nouns the definition should or should not include. Here, we will adopt the definition given by The Webster's Third New International Dictionary of English Language, 1976. It tends to be the most widely accepted. According to it, a proper noun is:

> " a noun that designates a particular being or thing, does not take a limiting modifier, and is usually capitalised in English".

To illustrate the difficulties involved in the definition of proper names, we take the discussion as to whether temporal names like names of the days of the week, months or recurrent festive days are to be seen as proper names. Some argue that as there is one Monday each week, one month of June and one Good Friday each year, they do not really designate unique temporal events, but rather categories of events, and therefore these are not true proper names. It is claimed that support for this view can be found in the fact that in other languages, such as in French and in Portuguese for instance, these nouns are not initially capitalised. There are many other similar difficult issues surrounding the definition of proper names. The point of view taken here is that proper names are names of unique beings or things.

Proper names are constantly being created, so that they can attend to people's need for referring to new institutions, shops, buildings, products that are introduced into the market and so on. In this respect a remarkable feature that distinguishes the category of proper names from that of common nouns is the demands it puts on the cognitive system: whereas the rate of acquisition slows down for common nouns in adult life, proper names have still to be learnt throughout life (Valentine, T., Brennen, T., & Bredart, S., 1996).

Proper names encountered by people for the first time in a specific context, are what is called here a *new proper name*. Some are old names that have been around for quite a long time as part of one culture. Then, for some reason they break their cultural barriers and are learned by a bigger crowd of people generally from another culture. This is the case, for example, of the Russian name "Yeltsin". However, there are also the genuinely new brand names. To use the classification system proposed by Valentine et. al. (1996), these are of three categories: novel, derivational and novel combination names. *Novel* names consist of novel phoneme strings such as *Kodak* and *Exxon*. These are generally brand names or names of fictional characters generally found in films and books. *Derivational* names are those where the root of the word carries meaningful associations to the referent, but the name as a whole is new. For instance, the name of the administrative capital of Brasil is Brasília. *Novel combination* names differ from the two categories above in that they are old words or new combinations of old words. The words *apple* and *Macintosh* were familiar to British people, before they were put together to make the name of a computer. Another interesting example of a brand name, composed of very familiar words is *Word for Windows*.

Phonetic symbolism is a controversial and also a fairly peripheral aspect of the pronunciation system, however, it is worth mentioning here in the brand name context. The idea is that the sound-meaning relationships in a spoken language are not arbitrary, so that individual phonemes may have consistent meanings associated with them. This is interesting in relation to the creation of new words, particularly for proper names, as the name can be manipulated to suit the image that is to be conveyed. Phonetic symbolism has been shown for many aspects of language, even cross-linguistically. An interesting approach to phonetic symbolism was taken by Rubin, Stoltzfus, & Wall. (1991). They were interested in whether word domains have characteristic surface forms and performed two experiments. In the first one, a group of subjects were asked to list all radioactive elements, all the types of pasta, all the names of brands for painkillers and all the names of laundry detergents that they could remember. In the second experiment, a different set of subjects were asked to generate one new exemplar for each of the four categories. They concluded that subjects do indeed rely on prototypical surface forms when generating a new word from a particular category and when assigning new exemplars to given categories. For instance, there was much inter-subject agreement that radioactive chemical elements ended in 'ium'. However, their interpretation of these results is disputable. They have shown that subjects rely on prototypical information, but the results of their experiments do not go further to show unequivocally that this prototypical information is surface form.

4.1.1 - Proper Names in various disciplines

It is only recently that cognitive psychologists have started exploring the issues surrounding the subject of naming. However, the act of naming is one of the most basic assets of society. In this section we present some of the disciplines that carry out research on the topic of naming. An example, of the importance of the topic can be found in the fact that in all societies there are either explicit or implicit laws that regulate the act of naming new-born babies. The invention of proper names is a consequence of the act of naming.

First, the domain of law governs the protection of personal and commercial proper names. However, the law varies across cultures. There are some countries, for instance, where personal names are protected against commercial use (e.g., Belgium). Brand names are regarded as a subcategory of proper names. For a different point of view refer to Valentine et. al. (1996). Brand names can become generic, that is, they have been so widely and frequently used that they enter languages as words in their own right. A genericised brand name can be declared a public property with its creator having no more rights over it.

Proper names have also being taken up as a theme in social psychology studies. Social psychologists interested in the effect of name attractiveness on social judgements, have generated a great deal of research which investigates the consequence of first name stereotyping. One of the aspects studied by Kamin (1958) and O'Sullivan et. al. (1988) was on the influence it exerts in the electability of political candidates. It was found that candidates' surnames can indeed affect their electability, especially when voters are called upon to choose between candidates they have little information about.

Anthropologists are concerned with the many features that differ across cultures in the practice of naming and in the use of proper names. Many of the studies carried out investigated the diversity of practices related to issues such as: the moment of name giving, the continuity of naming, the plurality of names etc. However, in spite of all this diversity, anthropologists were also able to expose some principles that seem to be present in the act of naming in all societies. The universal function of a name seems to be a social classifying device that enables people to be classified in terms of parental, social, ethnic or geographic groups.

The semantic status of a proper name has long been a matter of prolific studies by philosophers of language and linguists. In brief, this has chiefly taken the form of a debate between on the one hand the description theories of reference (Frege, 1892; Kripke, 1980) and on the other hand the theory of direct reference (Russell, 1905; Searle, 1969). According to description theories of reference, a proper name can designate a person (or another unique entity) only via intermediate descriptive properties. By contrast, the theory of direct reference prescribes that proper names are directly linked to their bearer without intermediate descriptive properties. Philosophical theories of reference have also influenced cognitive psychology and neuropsychology. Cognitive psychologists address the question of whether the link between name and bearer is direct or mediated, by employing the mental representation of individuals, their properties and their names.

4.1.2 - Proper names and cognitive psychology

There is a considerable literature in cognitive psychology related to proper names. A full description is beyond the remit of this study, so we address only the most prominent issues here. This section is divided in two topics: the first deals with the literature regarding experiments with normal (non-impaired) subjects, the second examines neuropsychological studies of failures in retrieval and production of proper names.

4.1.2.1 - The experimental literature

Proper names are particularly difficult to recall and are particularly vulnerable to the effects of ageing (Cohen & Burke, 1993) and brain-injury (Warrington & Clegg, 1993; McKenna & Warrington, 1980; Lucchelli and De Renzi, 1992). These findings have prompted cognitive psychologists interest in the processing of proper names (Bolla et. al., 1991; Reason & Lucas, 1984; Young et. al., 1985). The claim that proper names are frequently more difficult to retrieve than other categories of words has been experimentally confirmed by Cohen and Faulkner (1986). They presented people with unfamiliar faces and taught them various facts about the people in the photographs; for instance, their profession, their hobby and their name. Cohen and Faulkner found that names were indeed more difficult to recall, than other pieces of information. For example, proper names were responsible for a larger percentage of the items incorrectly recalled. Additionally, there was an effect of age on recall: the older subjects (mean age 71) had more difficulties in the recalling of names than the other two age groups (mean ages 31 and 47 respectively).

One of the driving questions behind most of the contemporary cognitive psychology research on proper names is "why are proper names so difficult to recall in comparison to other category of words ?" To answer this question McWeeny et. al. (1987) developed the *association of information to unfamiliar faces* paradigm in which subjects are asked to learn the names and occupations belonging to unfamiliar faces. The same word was presented sometimes as a name and sometimes as an occupation (e.g., Porter-porter, Tailor-tailor). This means that imageability, frequency, distinctiveness of names and meaningfulness were equated for the name recall condition and for the occupation recall condition. The striking result was that when subjects recall the name of a face they consistently recall the person's profession. Conversely, they often could recall the profession, but not the name of the person. This was true even if the proper name plus the profession were the same (e.g. Baker-baker). This result has been termed the "the Baker-baker" paradox. The data has clearly shown that the difficulty in recalling people's names cannot be attributed to any features of the names per se (e.g. imageability, frequency or distinctiveness of names), because there is a recall disadvantage for proper names even when the same orthography and phonology were involved.

4.1.2.1.1 - The tip-of-the-tongue phenomenon

Most of the time, we are proficient in retrieving words from the vast pool that is the mental lexicon. Other times, the feeling of imminent recall is present, but for some reason we are unable to retrieve the desired lexical item. This annoying type of experience is a familiar one and became known as the tip-of-the-tongue (TOT) phenomenon (James,1893; Boilinger,1961). Brown and McNeill (1966) were the first to study TOT states experimentally. They examined the types of phonological information available when someone is in a TOT state. Subjects were given definitions of relatively uncommon words, such as the *sextant* example below:

> "a navigational instrument used in measuring angular distances, especially the altitude of sun, moon, and stars at sea."

The subjects had to retrieve the name of the object. Some subjects knew the instrument's name immediately; others could not remember it at all. But some few went into a TOT state. They were asked to guess the initial letter and the number of syllables, to mention the words that had come to mind, and so on. For the above example, subjects tended to guess /s/ as the initial phoneme and two as the number of syllables, and sound-related words like *secant* and *sextet* had come to mind.

The tip-of-the tongue (TOT) phenomenon has more recently become one of the means used to investigate adults difficulties in recalling proper names. Two approaches are normally used, although with neither of them a high degree of certainty can be achieved. Due to their spontaneous nature, one way is the development of techniques to help in evoking TOTs in a laboratory controlled environment. Yarmey (1973) pioneered the study of TOT states in adults' recall of proper names eliciting them by presenting subjects with photographs of the faces of famous people which they had to try to name. Another way is to record TOTs as they spontaneously occur during daily activities (diary studies). The use of this type of naturalistic data, such as diary recordings is popular in the study of retrieval blocks in the production of proper names (Reason and Lucas, 1984; Cohen and Faulkner, 1986; Burke et. al., 1988, 1991). The above studies have produced a plethora of interesting findings. For example, age related differences were found: the occurrence of TOTS involving proper names is higher for older subjects than for younger ones (Cohen and Faulkner, 1986).

In Brennen et. al. (1990) TOTs were induced by reading out definitions of famous landmarks and description of famous people to the subjects. The ability of different cues to resolve TOTs was compared and it was found that the initial letter(s) of the name(s) of the target reliably resolved the TOTs, whereas a picture of the target did not resolve more TOTs than mere repetition of the original TOT-inducing information.

Burke et. al. (1991) have made use of both methods of obtaining TOTS above (naturally occurring and induced TOTs). They used recordings of naturally occurring TOTs and also attempted to provoke TOTs by reading general knowledge questions to subjects that had common nouns and proper names as their answers. According to Valentine et. al. (1996) this is the only experimental study that has been reported in the literature comparing the incidence of proper names blocks with that of common nouns blocks. This combined study suggested that ageing has a far greater effect on the retrieval of proper names than the retrieval of common nouns.

4.1.2.2 - The neuropsychological literature

Before we focus on the cognitive neuropsychological processing of proper names, we will examine some broader aspects of this branch of psycholinguistics. Cognitive neuropsychology is defined as follows:

> "Cognitive neuropsychology studies the underlying mechanisms of the psychological processes that are the basis of our mental life - thinking, reading, speaking, recognising, remembering - through the effects of brain injury. Its first aim is to relate the patterns of cognitive performance in brain-injured patients to psychological operations that are necessary for normal cognitive function. The second is to actually draw conclusions about normal cognitive processes from observation of the effects of brain injury.(Springer and Deutsch, 1993)."

One of cognitive neuropsychology's main contributions has been the demonstration of the independence of specific types of information processing. Broca (1861) for example, illustrates the phenomenon of dissociation between speech production and comprehension, in his description of patients who were unable to talk but who could understand language. Also the phenomenon of impaired comprehension with the preservation of speech production (double dissociation), has been observed by Wernicke in some of his patients (1874). Later we will see that these patterns of dissociation and double dissociation are important clues in the investigation of proper name organisation.

More specific disorders are also reported in the neuropsychology literature: these include the category-specific disorders. The patterns of deficits presented by these cases have been interpreted as indicating *unavailability* of the semantic features specific to the elements belonging to a certain conceptual domain, for example, names of geographical places. Sometimes researchers have observed *preservation* of the semantic features belonging to just one specific domain, for instance, body parts and loss of the semantic features pertaining to all other domains. A particularly specific naming deficit is described by Hart et.al. (1985). Their patient, MD, was able to satisfactorily name stimuli from all tested categories except fruits and vegetables. In a categorisation task, she was unable to categorise pictures of fruits and vegetables, implying a semantic deficit affecting only the representation of these items.

Proper names are a particular case of a conceptual domain affected by brain injured disorders. The overall findings offer a complicated picture, where we may find opposite patterns of dissociation: either a specific name finding impairment in the absence of any other evidence of language disability, or a preserved island of proper names comprehension, in the context of a profound aphasic disorder.

We now turn to neuropsychological studies involving the impairment and preservation of the category of proper names. The importance of these studies, as will be seen, amounts to the suggestion that proper names possess an independent status in comparison with common names.

To illustrate the above, next we report a few of the many case studies where a deficit or preservation of the proper names category is involved. One of the patterns that can be found in these cases is for example, that some patients present deficits regarding only a sub-category of proper names, like people's names.

Mechthild, et. al. (1997) report a case, where the patient CU presented a cross-modal anomia for people's names. That is, CU failed to name people in a verbal and several visual naming tasks. However, she was perfect in naming pictures or generating common and proper names of other semantic categories and was also unimpaired in the recognition of objects and faces.

Fery et. al. (1995) present the case of OV, a French-speaking, 63 year old woman who had suffered an aneurysm and had acquired the same type of anomia described above, that is specific for people's names. OV has no impairment for other types of proper-names or for common names. Her deficit with people's names was equally present both in face-naming and in naming from definition and was not affected by the descriptiveness of the label born by the individuals.

There are also reports of patients who exhibit impaired production for several types of proper names in the absence of word finding problems for common names. Semenza and Zettin (1988, 1989) report the case of two patients who had difficulties in retrieving names of people and also names of geographical sites such as rivers, mountains and towns, while able to provide correct and often precise information about the items to be named. These naming difficulties were observed in visual confrontation as well as in naming upon definition.

The number of reports of deficits that presented the reverse pattern, that is, the preservation of the category of proper-names in relation to that of common nouns, were rarer until recently. One of the first cases to be described was that reported by McKenna and Warrington (1978). They reported the case of FC, who could name only countries and could also match countries to appropriate objects, demonstrating access to semantic information about this domain. Nevertheless, in all other tested domains, he could neither name stimuli, nor was he able to demonstrate evidence of semantic access.

Warrington and McCarthy (1987), report a patient (YOT) who showed preserved comprehension of proper names along with impairment of comprehension of common nouns; but in this case the impairment for common nouns was category specific. In general, YOT's comprehension of objects was significantly more impaired than her comprehension of foods and living things. Particularly she showed a preserved ability to point out written proper names in an array which matched a spoken proper name and preserved ability to match famous people's names, countries and cities. However, she was severely impaired in matching common surnames, boys' names and girls' names. YOT also showed preserved ability to match a spoken word to the appropriate picture in an array for the categories of famous people, famous buildings and countries.

More recently, Semenza and Sgaramella (1993) describe a patient, R.I whose extremely severe aphasia was restricted almost completely to production deficits. He showed a preservation of proper names production despite an otherwise deeply troubled linguistic production.

Finally, important insights into the processing of proper names can be drawn from comparison of production with comprehension deficits. Goodglass and Butters (1988) and Goodglass and Wingfield (1993), for example report on a double dissociation in comprehension of a class of common names (body parts) and a class of proper names (place names) in studies with a mixed group of aphasic patients (Broca's, anomics' Wernicke's aphasics etc). In both studies, Wernicke's aphasics and global aphasic patients show relative preservation of comprehension of geographical place names compared to body parts, but anomic aphasics showed better comprehension of body parts than geographical places. The relevance of studying these type of dissociations involving proper name anomia and comprehension resides in the prospect they bring of unveiling some of the brain functioning mechanisms. As put by Goodglass and Wingfield, themselves;

> "... [they] may serve as an important stimulant for resolving these and other long-standing questions in cognitive neuropsychology"

More and more neuropsychological studies on language impairment are being reported. The type of findings emerging from these studies suggests the organisation of language in the brain to be a much more complex and sophisticated picture than had been anticipated just a few years ago. This development had the effect of leading many cognitive psychologists to replace the box-and-arrow types of models explanation of language with mechanisms that are more grounded in terms of biological principles. One such mechanism is that offered by Pulvermüller in his most recent paper (Pulvermüller, in press). Starting from the fact that human language production is caused by neuronal activity and any speech signal necessarily activates neurons in the brain of listeners when being perceived, he develops a theory of how language works in the brain. At the heart of his explanation is the proposal that areas of the brain are specialised in different activities, however these areas are interconnected through a widely distributed network. Not long ago, it was believed that the hemisphere that dealt with language was typically the left one. Therefore, left hemisphere activation was to be expected independently of what type of word was being accessed. Pulvermüller, argues that brain imaging techniques have shown that different patterns of activation occur, for example, in the case of function and concrete words. Whereas for function words the main activation in the brain occurs in the language areas pertaining to the left hemisphere, for concrete words activations also happen in areas that are not specialised in language such as the visual cortex. He concludes, that this is so because concrete words, such as "chair" evoke visual associations that are processed in both hemispheres. Pulvermüller's assertion that other areas of the brain than those directly responsible for language processing are activated during the access of words is certainly intriguing. More interesting still is the link he proposes between the different modalities that are involved in the learning of a word and the increase of activity in multiple areas of the brain that are not usually associated with language. This type of mechanism suits best as an explanation for the type of phenomena described above concerning neuropsychological accidents that involve category-specific impairments. We hypothesise that categor-specific impairments of names are due to the special nature of proper names, for example, whereas common names refer to a class of objects (type-reference) proper names usually refer to single objects (token-reference; Jackendoff, 1983). The relationship between the word and what it represents differs for common names and proper names in that it is more "rigid" in the latter. The word "Taylor", for example, is used to refer to the individual "Taylor" in any utterance or text. By contrast, the word "chair" can be replaced by "something to sit on" or even by its Portuguese equivalent "cadeira" according to the situation (Kripke, 1982). Finally, most proper names, particularly ordinary person names, do not possess a meaning understood as a set of properties, like common names do. The type of information that is associated with proper names, especially people's names has social and emotional connotations and this has to involve massive amounts of information that probably is more widely distributed between the two brain hemispheres if compared to information that concers common words. There are two possible implications of this view, in terms of the vulnerability of proper names. First, proper names could become more vulnerable to brain damage, since bits of information pertaining to them are now more widely distributed in the cortex. Therefore, damage to any area of the cortex can potentially bring about proper name impairment. The inability to exactly pin-point areas responsible for category-specific deficits in brain damage patients is actually one of the sources of disapointment in neuroscience. A second outcome is that proper names by having more content addressable storage become less vulnerable to damage. Unfortunately, there is no decisive neuropsychological data available up to now capable of undoing this impasse. In addition, the means of eliciting name-judgments will also affect apparent level of impairment

In summary, the landscape presented by neuropsychological studies of the language function system is a complex one. Even in the specific case of proper names, double dissociation patterns, modality dependent symptoms, impairments that sometimes affect the whole category of proper names and other times affect only some of its sub-categories, islands of preservation, etc., serve as hints of the elaborated brain's pathways. However, these studies do suggest the possibility of proper names holding a special status and possessing a particular type of processing that differ from that of common words. Contemporary models explaining language mechanisms that are based in biological aspects seem to be the best suited one to cope with the type of findings presented by neuropsychological reports.

4.1.2.3 - The set size of plausible phonologies (SSPP)

Brennen (1993) stated that there is a wider range of phonologies that are plausible and acceptable for people's names than for common nouns. He means that, during their lives people encounter and learn a host of persons' names, such as "Depardieu", "Yeltsin", "Miranda" that entertain different phonologies. By contrast, the learning of new common words, such as names of professions belongs to a more restricted domain, perhaps a profession such as "cognitive scientist" would have not been encountered by some people before. He observes that a consequence for adults is that the rate of acquisition of new common nouns slows down quite considerably, while the learning of novel people's names continues steadily throughout our lives. Furthermore, it is not only the learning of people's names that contributes to the increase in the range of plausible phonologies, but also the learning of new brand names as well (e.g., SUBARU, HONDA). This results from massive advertising campaigns that are all the time being launched in the media. To illustrate his observation, he asks us to imagine that we were told that someone is called Mr. Dreaner. We might perceive this as an unusual name, but have no problems in accepting it as a person's name. However, if we were told that a person works as a *dreaner*, than probably we would think that we have not heard it correctly. The reason for this, he argues, is that our experience tell us that many more phonologies are plausible for people's names as compared to common names: we frequently encounter people's names that consist of unfamiliar phonologies.

Before continuing let us stop and make a criticism of Brennen's choice of term "plausible phonologies". Perhaps, he would have done better in choosing a term such as "alternative phonologies" or "feasible phonologies", that express in a more explicit manner, the actually common sense ideas that are behind his theory. By "plausible phonologies" Brennen means a repertoire of some crude knowledge about the phonology of different languages, for example, the vague awareness of some of the different segmentation strategies that can be used in German, Greek or Chinese, or also some knowledge about phonotatics rules of other languages that do not apply to English. This type of knowledge, he argues, is acquired simply by our everyday exposure to a host of foreign proper names that arrive via different sources in virtue of the technological advances in telecommunications. He speculates that the consequence of that is that we take a more relaxed approach in terms of phonological rules when dealing with the pronunciation of lexical items that belong to the category of proper names.

Brennen's explanation for why recall of proper names is more difficult than that of common nouns depends on the assumption that the access to a word's phonology is a gradual process, i.e., it is accessed piecemeal. Support for this assumption comes from findings that when subjects are in TOT states, before full phonology is available, they show partial knowledge of word targets, such as knowledge of the initial letter, number of syllables it contains and (in English, at least) its stress pattern.

Further, Brennen asserts that different domains of words have different 'new exemplar' rates. People's names for instance, is a domain with a larger rate of new exemplars than, say, names of professions. He also points out that in a domain with a large new-exemplar rate, neighbouring (perhaps unfamiliar) phonologies are credible alternative pronunciations for an exemplar of that domain. The consequence is that a word from a large new-exemplar rate domain will have to be specified more accurately in order to distinguish it from potential alternatives. This is not the case for words from domains with small new exemplar rates.

According to what has been said above, the domain (or category) of people's names has a larger rate of new exemplars than the domain formed by common nouns. So the number of plausible phonologies for a given name is larger than that for a common noun. Therefore, it should be easier to guess the complete phonology of a target word from incomplete phonology if that word is a common noun than if it is a person's name. In other words, partial phonology accessed during recall of a person's name is less specifying of the target than is partial phonology during recall of a common noun.

Brennen argues that the set size of plausible phonologies (SSPP) plays the role of a top-down constraint in the recall of names. By that he means that the SSPP works by providing important information that helps in restricting the extent of the search process for the name's full phonology. He offers *high familiarity* as another example of a top down constraint. *High familiarity* restricts the size of the set of words to be searched for the correct phonology, by allowing only the pronunciation of highly familiar words to contribute to the construction of the articulation of the words to be read aloud.

The words for which the SSPP constraint will be useful for specifying the phonology will be those that come from domains with low new exemplar rates, e.g., names of professions. For such items, the full string of phonemes can be generated from a few phonemes, because these first-recalled phonemes will tend to uniquely specify a member of the target word domain. This, in concert with the knowledge that there are very few new exemplars in this domain, will allow the generation of the desired phonology with a high degree of certainty. For word domains with high new exemplar rates, this top-down constraint is simply not restrictive enough. For people's names, as argued earlier, very many phonologies are plausible, and so extrapolating from some partial phonology to a particular phoneme string is less likely to result in the desired phonology.

Brennen claims that the SSPP mechanism can account for the results of two experiments. The first is the Baker-baker paradox (McWeeny et. al., 1987) referred to before. He argues that in the recall attempt, the utility of a partial phonology that is already accessed in specifying the rest of the word's phonology will differ according to whether it is a surname or a profession that is being recalled. For surnames, the constraint of the SSPP will be very weak, and so before a surname can be recalled all of its phonology must be specified. In contrast, for the name of a profession a small amount of partial phonology can reliably specify a particular lexical entry.

The second experimental evidence is that of Cohen (1990). Subjects were asked to associate a profession, a surname, and a possession to unfamiliar faces. Cohen noticed that when the possession was a real word, e.g., 'boat', subjects recalled the surnames less often then the other two pieces of information. However, when the possession was a nonword, e.g., 'blick', surnames and possessions were recalled less often than professions. The explanation given by Cohen for why names are not better recalled than nonwords, is that both categories are processed as meaningless entities. However, the data is also consistent with the plausible phonologies hypothesis. The set size of plausible phonologies will be large for both nonwords and for names of people. The partial access to phonologies will be of negligible utility in recalling a nonword, as in recalling the name of a person. The SSPP for professions on the other hand is much smaller, making a partial phonology already accessed more powerful in the recall process. However, it should be considered that if there are any phonological constraints on a person's name, then one would expect that person's name would be better recalled than nonwords. There should not be any phonological constraints for nonwords. Nevertheless, as Brennen points out, it might require a more powerful paradigm in order to observe this effect. However, he does not make any suggestions as to what type of paradigm this might be. We claim to have found this paradigm, that will be presented later in the experimental section. Also, he does not venture any suggestions of what type of mechanism could account for this theory. Later, in our general discussion we discuss a type of mechanism that can account for our experimental data that was interpreted on the light of Brennen's "plausible phonology" theory.

4.1.3 - Models of face recognition

From what we have seen so far, the access to names is independent of, and dissociated from, access to other semantic information. There is also some evidence that the retrieval of proper names is more difficult than that of other categories of words. Earlier, in section 4.1.2.2, I have argued in favour of a more biologically based mechanism as an explanation for proper names processing. I have criticised the neat approach that is adopted by the box-and-arrow models as an explanation to such complex phenomena as the category-specific impairments that were presented.

It must be clear however, that in doing so I had no intention of denying the important role played by this type of model in the progress of psychological theories. As mentioned earlier, interest in the area of proper names has partly been renewed by methodological and theoretical advances in the domain of face recognition and the most relevant models in this area are not biologically based models. Next, we present a short description of three models of face recognition, which claim to be able to account for the findings above. Two of these models are theoretical and the third one is computational.

4.1.3.1 - The Bruce and Young model (Bruce & Young, 1986)

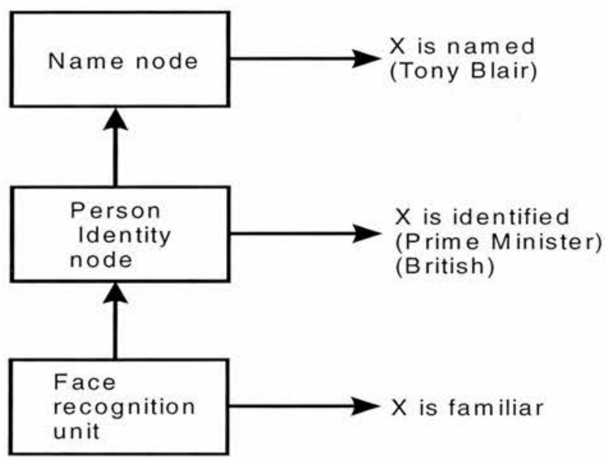

Fig. (4.1) - Successive stages in the recognition of a person: A simplified version of the Bruce & Young model. (Stanhope & Cohen, 1993)

A number of studies have shown that the recognition of familiar faces is analogous to the recognition of objects. It has also been observed that the recognition of faces shows similar effects to those observed in visual word recognition. The framework of the model to be described next, as well as the concept of face recognition units used in the model, were derived from the logogen model of word recognition (Morton, 1969, 1979) and also from models of object recognition (e.g., Warren and Morton, 1982).

According to this model, the recognition of a person is achieved by serial access to a set of hierarchically arranged components. The first component is a face recognition unit (FRU). It is postulated that for every known face, there is a FRU representation stored in the brain. Perception of a known face activates a stored face recognition unit (FRU) and the face is then recognised as familiar. The FRU then activates a person identity node (PIN) where semantic information about a person is stored. This information becomes available once the PIN is activated. Names are stored in a terminal node which is the final stage of the sequence and can only be accessed via the PIN: there is no direct link from the face to the name. This serial access model can therefore account for the fact that names are always accessed last in comparison to the semantic information available. It can also account for the fact that name retrieval can be independently impaired. It is also consistent with the finding that person identity information is often available when the name cannot be accessed but that the contrary case, retrieving the name without recalling any person identity information is extremely rare.

Their model specifies the relationship between different aspects and stages of face processing. It deals with many of the possible types of processing that one can undertake upon seeing a face (e.g. judging facial expressions, lip-reading etc.) It comprises information of an input code, activation of a face recognition unit, access to semantic information including a person's biographical and contextual information and finally, access to the person's name.

4.1.3.2 - The Valentine et. al. model (Valentine et. al., 1991)

The Valentine et. al. (1991)[1] model is an extension of Bruce and Young's model (see Fig.4.1) and its use here is two-fold. First, it is being used as an illustration of how proper name studies are interrelated to those of face

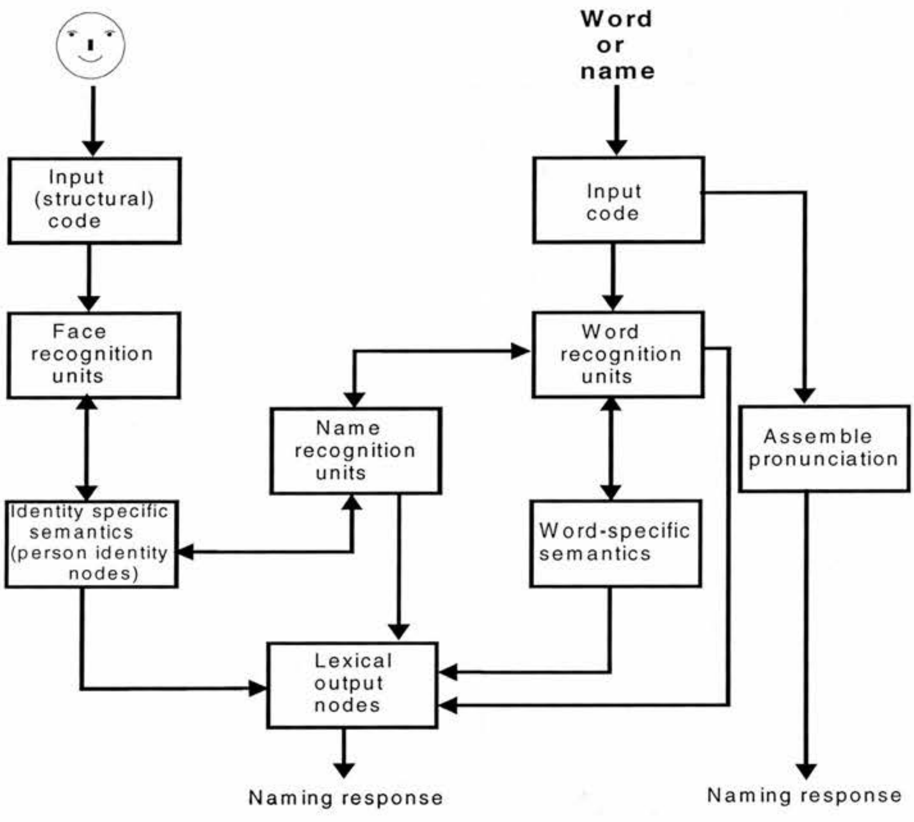

Fig. (4.2) - A functional model of face, name and word recognition proposed by Valentine et.al.(1991). Note that name mediating between word recognition units and identity specific semantic nodes lies the name recognition units.

recognition. Second, one of the underlying assumptions of this model is that proper names are a special category of words that must be stored separately from all the other common words. See Fig.(4.2).

[1] A later, more sophisticated version of the model, has been put forward to accommodate more recent neuropsychological findings, as well, as later theories and experimental results (Valentine et.al., 1996). However, it will not be described here, as these modifications are not relevant in the present context.

The model assumes a separate access route for proper names (at least names of people), and for common nouns, to the phonological output codes. It is suggested that name recognition units (NRUs), mediate between the word recognition system and access to identity-specific semantic information about individuals (see Fig. 4.2). The output of word recognition units (WRUs) which represent names connects to NRUs. The input to NRUs could be first or surnames alone, initial and surname or full names. There is a WRU for every familiar word (or name) and there is a NRU for every known individual. The recognition of people's names is mediated by a set of name recognition units (NRUs). Phonological output codes can be accessed directly from name recognition units. This route is analogous to the direct route from word recognition units to phonological output codes.

The most interesting feature in this model is the concept of "name recognition units", where names are stored separated from their semantics. It is also the most debatable one. As it will be seen, the connectionist network described next was put forward as an implementation of the Valentine's model, however it departs from the model in some fundamental respects, such as that of having names and specific semantics stored in the same pool.

4.1.3.3 - The IAC model (Burton & Bruce, 1993)

The Burton and Bruce (1993) model is based on an earlier Interactive Activation and Competition (IAC) simulation of normal and impaired face processing and naming (Burton et. al., 1990; Burton & Bruce, 1992). There is a whole class of models known as interactive activation models. In these models, activation is input to a number of units. It is then passed along excitatory links to connected units. The activations of units change over time, and this change is ideally continuous.

This process is modelled by a number of cycles, in each of which activation levels are updated. After many cycles, the activations of the units tend to stabilise. The Burton and Bruce model uses localised representation, where each concept is represented by one individual unit of the network. It is composed of six pools of units, three of which are remnants of the previously mentioned model (Burton et. al, 1990). These are face recognition units (FRUs), person identity nodes (PINs) and semantic information units (SIUs). The following pools of units were included in the 1993 model: the lexical output

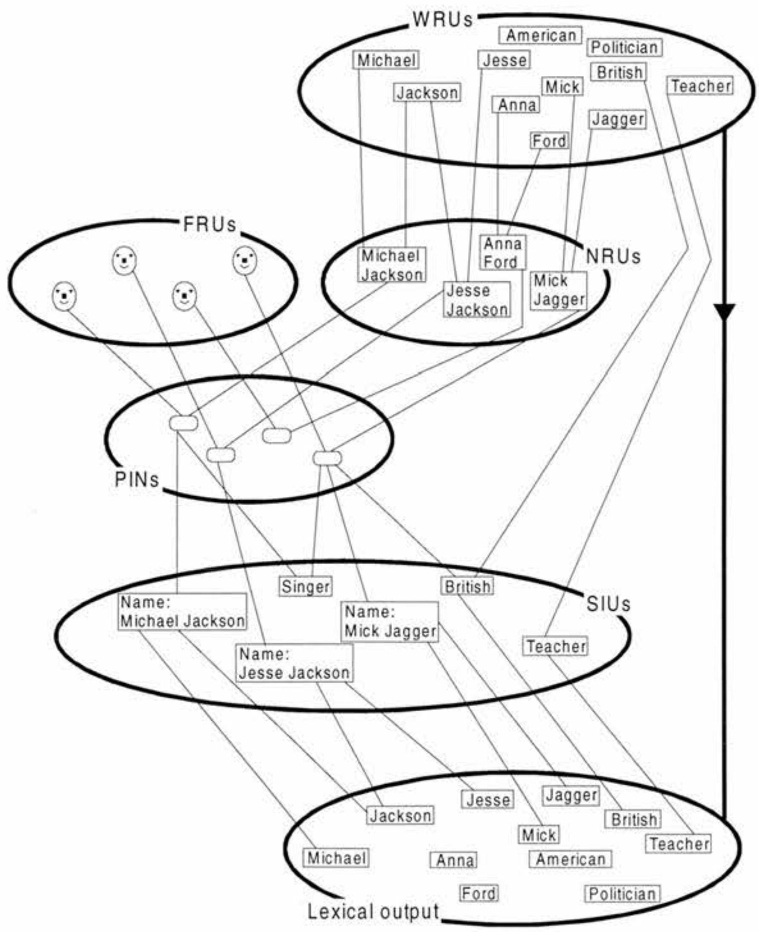

Fig. (4.3) - The architecture of Burton and Bruces's (1993) implementation of the Valentine et. al. framework using an interactive activation and competition model.

pool, the name recognition pool (NRUs) and finally the pool that represents the word recognition units (WRUs). The way all these units are connected can be seen from the illustration of the model in Fig.(4.3.). The previous models could only account for the retrieval of names. Enriched by the newly added pools, their new version is now able to simulate not only the retrieval but also the production of proper names. However, to enable this the identity-specific semantic information is stored at SIUs which can only be accessed via PINs, when a person is being recognised. Those WRUs that do not code names bypass the person recognition system, and instead are connected directly to SIUs - units coding their semantics. In doing so they depart from Valentine's approach and the assumption that names are stored separately from identity-specific semantics. According to the Valentine et. al. model, one of the reasons why proper names behave somewhat differently from common words is that they are not stored together with semantics. Burton and Bruce (1993), claim that proper names are harder to retrieve than other types of information because they are typically unique entities, i.e., there is only one Tony Blair, who is the Prime Minister of Britain, and this is what makes them harder to retrieve than other types of information.

4.2 - The English grapheme-phoneme relationship

Let us now turn our attention to the more general area of pronunciation and models of reading aloud. We begin with an introduction to the grapheme-phoneme relationship in English.

The basic unit of written language is the *grapheme*. This can be defined as a letter or combination of letters that represent phonemes. For example the word "thought" contains seven letters and three graphemes ("th", "ough", "t"), representing three phonemes. There is much more variability in the structure of written languages than there is in spoken languages. Whereas all spoken languages utilise a basic distinction between consonant and vowels, there is no such common thread to the world's written languages. The most familiar sorts of written language to the Western world, are alphabetic scripts, such as English. In alphabetic scripts, the basic unit represented by a grapheme is essentially a phoneme. However, the relationship between the written and the spoken forms varies across languages. Alphabetic orthographies may be characterised by depth - i.e., the extent to which they are phonetic representations of speech. Some languages, such as the Serbo-Croatian, are said to have a *shallow orthography* (Katz & Feldman, 1981), meaning that the relationship between grapheme and phoneme in this type of language is consistent. The orthography generally represents the phonemic level of the spoken language accurately. Because in a shallow orthography the grapheme-phoneme relationship is invariant, guessing the pronunciation of a novel word, for example, is not a problem and in fact it could be accurately achieved by making use of a set of rules.

Other languages, like English are classified as having a *deep orthography*. It does not have a transparent correspondence between spelling and sound. There is not a one-to-one mapping between orthography and phonology. For example, in English a phoneme can be realised by many different graphemes (e.g. the long /aɪ/ phoneme in the written words "mine", "pie" "my"). Conversely, a grapheme can also be realised as many different phonemes (e.g. the letter "a" in the words "fate", "pat" and "wash"). To put it in another way, the correspondence between the written and spoken codes in English gives rise to many inconsistent correspondences (e.g. the block -AVE is usually pronounced as in GAVE, SAVE and CAVE, but there is also HAVE) or wholly arbitrary correspondences such as (-OLO in COLONEL, -PS in CORPS). These inconsistencies derive from several sources. According for example, to the theory of generative phonology (Chomsky and Halle, 1968), in English the consistency of the pronunciation is traded-off for simultaneously encoding phonological and morphological information. The lack of a direct correspondence between letters and phonemes in words with apparently irregular pronunciations such as GRATEFUL and DIVINE is explained by the preservation in their written forms, of information about morphological relations among them (GRATEFUL - GRATITUDE; DIVINE - DIVINITY). Further sources of inconsistency include the changing pattern of word spelling over time, lexical borrowing from other languages, periodic spelling reforms and also historical accidents.

According to some estimates, 20% of all word types in English violate regularisation rules (Hanna & Hanna, 1959). Henderson (1982) points out that the most common 3000 words in English, as many as 21% break the rules proposed in Wijik (1966). However, for competent English speakers these inconsistencies seem not to pose any hindrance.

4.3 - Models of reading

Any model of reading in English has to take into account that, as a consequence of its deep orthography, words are classified into two broad categories - *regular* and *irregular* (exceptions). Regular words are those for which graphemes map onto phonemes in a totally regular way, i.e., no special knowledge is needed to know how to pronounce them. The mapping of graphemes to phonemes in irregular words is that of many to many, in other words, more than one grapheme can represent one phoneme plus more than one phoneme can represent one grapheme. Consequently, irregular words do not follow the expected pattern of spelling-to-sound correspondences (e.g., "pint", "steak"). Also, there are different types of irregular words. Modelling has also to take into account the behaviour of novel words or pronounceable nonwords (pseudowords). The main lines of research in this area will be outlined here. The two main theoretical lines in word pronunciation can be described by the dual route model and the analogy model. Following these, two connectionist models of pronunciation will also be discussed.

4.3.1 - The dual route model

The dual route hypothesis posits that there is both a direct visual route to the lexicon and an indirect phonological route mediated by the grapheme-phoneme correspondence(GPC) rules. The direct visual route can be understood as a look up dictionary procedure that explains the reader's ability to read quickly the words they are already acquainted with, i.e., already represented in their mental lexicon. In the lexical look up, the reader gains access to a lexical entry and extracts from this entry the information necessary to generate an articulation of the word corresponding to that entry. This succeeds for regular and exception words.

The phonological route, on the other hand, admits the possibility that a reader possesses an internal system of grapheme-phoneme correspondences (GPCs), and that a string of letters is decomposed into its constituent graphemes, after which the GPC system is used to assign a phoneme to each grapheme. This succeeds for regular words and for pronounceable nonwords. It fails for exception words, because the GPC procedure yields incorrect pronunciations for exception words. As neither of these routes alone can adequately account for reading performance, it is assumed that we possess both. More recently, dual-route theorists have conceptualised reading as a race between these two routes. When a word is seen, both routes start processing it. Most of the time the direct route is much faster, so this will usually "win" the race.

The main support for the dual route model comes from neuropsychological data regarding the two types of acquired dyslexia, surface and phonological dyslexia (Marshall & Newcombe, 1973; Patterson & Marcel, 1977; Shallice and Warrington, 1975). Surface dyslexia is a form of acquired dyslexia in which the reading aloud of nonwords and regular words is selectively preserved relative to the reading aloud of exception words, and exception words are also often read as the GPC rules specify. For example, *glove* may be read as if it rhymed with *cove*, or *flood* as if it rhymed with *mood*; such responses are referred to as regularisation errors (Marshall & Newcombe, 1973; McCarthy and Warrington, 1986). Phonological dyslexia on the other hand is a form of reading impairment in which the reading aloud of nonwords is selectively impaired relative to the reading aloud of words (Beauvois & Derouesne, 1979; Funnell, 1983). It is argued that the two basic categories of patients who acquire reading disabilities after brain injuries seem to demonstrate the independent existence of visual and phonological mechanisms of reading. The way the dual route model deals with these data is to propose that the skilled reader possesses a reading system consisting of a number of separate modules such as a letter-identification system, a visual word recognition system and so on. Some of these modules belong to the lexical route for reading and others belong to the nonlexical route. So, the impairment of one of these modules can impair the reading of one type of string but not others. *Deep* or *phonological* dyslexics pronounce and recognise familiar words but are unable to assign any reasonable pronunciation to novel letter strings. These patients apparently have an intact visual access mechanism with a loss of phonological rules. *Surface* dyslexics, however, pronounce words and nonwords equally poorly. Visual recognition of words is no longer possible, but phonological recognition sometimes succeeds when the phonological translation of a printed letter string "sounds like a word". This syndrome is interpreted as a loss of direct visual access with partial damage to the phonological mechanism.

4.3.2 - The activation and synthesis model

Glushko (1979, 1981) challenged the plausibility of the dual-route mechanisms and put forward an entirely new view of how the task of reading aloud would be accomplished. He proposed that the pronunciation of a string of letters emerges from the activation of an entire neighbourhood of words that share with it orthographic features that are activated in memory by the recognition of the individual letters composing the string. The pronunciation emerges through the co-ordination and synthesis of many partially activated phonological representations. According to this theory, familiar and unfamiliar words and regular and irregular words, are pronounced through a unitary process of activation and synthesis rather than by separate lexical and phonological mechanisms. Glushko claims that in this new framework, the classification of words into two absolute categories of regular and exceptions does not make sense. He revises the classification keeping intact the exception words, but makes changes in the regular category. According to his classification, "have" still is an exception word. If we take the word "hate", for example, it is still regarded as regular, because the words ending with "-ate" all attain the same pronunciation. Therefore, the pronunciation of words is always accomplished employing analogies. However, in the case of words like "gave" and "wave" they are regular words but of a different kind, being inconsistent words, since they are in the same neighbourhood as the word "have". They are presumably close enough to the exception word so that when they are pronounced, "have" might be activated and influence the process. In fact, just as his theory would predict, he has shown that nonwords which incorporate inconsistent letter-sound correspondences are indeed pronounced more slowly than consistent nonwords. He also showed that consistent real words like *blade* are pronounced faster than other familiar words like *brave* which contain inconsistently pronounced letter strings such as *bravo* and *breve* in their set of neighbours. This implies that analogies may contribute to, and sometimes retard, even the naming of familiar words.

4.3.3 - Connectionist models of reading

As seen above, a much discussed issue surrounding the models of reading aloud concerns the number of mechanisms involved in the adequate performance on both pronounceable nonwords and exception words. On the one hand there are those models based on the dual-route approach which argue for two separate mechanisms, one that applies rules of grapheme-phoneme correspondence and another that retrieves pronunciations specific to particular familiar words. On the other hand there are analogy-based models, that argue for pronunciation being achieved by the assemblance of parts of words forming the mental lexicon that are similar to each other (neighbours). A third alternative however, is embodied by the connectionist models of reading aloud.

In some ways, the analogy model can be regarded as a precursor to the connectionist models. For example, the analogy model assumes that only one route is needed for reading pronounceable nonwords as well as exception words aloud. However, in at least one important respect, the connectionist models depart quite radically from all the other models previously mentioned and that is in their use of distributed representation. In these models, there is no one-to-one correspondence between hidden units and lexical items; each word is represented by a pattern of activation over the input and output units. According to Seidenberg and McClelland for example, lexical memory does not consist of entries for individual words. Orthographic neighbours do not influence the pronunciation of a word directly at the time of processing. Instead, regularity effects in pronunciation derive from statistical regularities in the words of the training corpus - all the words we have learned - as implemented in the weight of the connections in the simulation. Lexical processing in this account involves the activation of information, and is not an all-or-none event.

Next, a description will be given of the essential features and performance of perhaps the two most prominent models of reading aloud in the connectionist literature. The first to be described is the Seidenberg and McClelland (1989) model and the second is the Plaut et. al. (1993) model. Later we will use the Plaut et. al. model in the discussion of our experimental results.

4.3.3.1 - Seidenberg and McClelland model (1989)

Seidenberg and McClelland (1989) developed a general framework for lexical processing where orthographic, phonological, and semantic information is represented in terms of distributed patterns of activity over separate groups of simple neuron-like processing units. Although strictly speaking the Seidenberg and McClelland (1989) framework has to be regarded as posing two routes, the semantic route proposed is a quite different thing from what was proposed before by the dual-route approach. Furthermore, this part of the model has never been implemented. The implemented part uses only one set of hidden units and only one process is used to name regular, exception and novel items. The architecture was implemented using the back-propagation algorithm which mapped orthographic representations (input) onto phonological representations (output) by means of a single hidden layer. It comprises a set of 400 orthographic units, 460 phonological units mediated by 200 hidden units. However, they did not implement the totality of their proposed framework. The implemented network contains no semantic or context information. The Fig. (4.4) below shows a scheme of the implemented part of the framework.

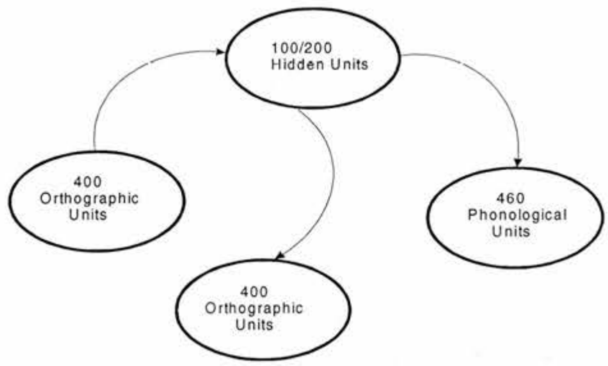

Fig. (4.4) - Implemented structure and number of units of Seidenberg & McClelland, 1989 model.

An important feature of the implementation is their choice of encoding for phonemes and graphemes. In representing a word's orthographic and phonological content, it is not sufficient to activate a unit for each of the letters or phonemes in the word because this would yield identical representations for pairs such as TOP and POT. It is also necessary to use some scheme that specifies the context in which the letter occurs. Their choice was a variant of Wilckelgren's (1969) triples scheme, known in the literature, after Rumelhart and McClelland (1986), as *wickelphones*. They are standard devices, used to represent position-specificity. The phoneme units are made context-sensitive by indicating the phonemes that precede and follow the phoneme of interest. Hence, each representational unit is sensitive to target phonemes and their immediate contexts. For example, the letter string "make" is treated as a set of letter triples #MA, MAK, AKE, and KE#, whereas the phoneme string /mAk/ is treated as a set of phoneme triples #mA, mAk, Ak#. The "#" indicates a blank space. Each wickelphone was encoded as a pattern of activation distributed over a set of representing phonetic features. A non-local representation was used: the graphemic representations were encoded as a pattern of activation across the orthographic units rather than corresponding directly to particular graphemes.

The number of words in the training set was 2897. The set was composed of uninflected monosyllabic words of at least three or more letters in the English language present in the Kucera and Francis (1967) word corpus. However, not all words were presented equally often. Words in the set were presented to the net a number of times proportionate to the log of their frequency of occurrence in the language. Each trial consisted of the presentation of a string of letters that was converted into the appropriate pattern of activation over the orthographic units, which then feed forward to the phonological units by way of the hidden units. Overall the net was trained on 150 000 learning trials to minimise the error scores.

After training, the network was used to explain performance in many different word-processing tasks. The general method was to present words to the trained net and compute the orthographic and phonological error scores. The orthographic errors were then used as an index of performance in lexical decision tasks and the phonological error scores were used as an index of performance in naming tasks.

In the ensuing debate over the properties of the model (e.g. Besner et. al., 1990; Seidenberg & McClelland, 1990), it became clear that although the model's simple architecture captured a surprising amount of observed word naming behaviour, it crucially failed to produce adequate nonword naming. Besner et.al.(1990) reported that, on nonword lists from Glushko (1979) and McCannan and Besner (1987), the model is correct only 59% and 51% of times, respectively, whereas skilled readers are 94% and 89% "correct". Although the model was generalising between the pronunciations it encountered in training, this generalisation was not sufficiently tailored to the detailed demands of naming behaviour. This was a serious empirical limitation in that it undermined its role in establishing a viable connectionist alternative to dual theories of reading.

4.3.3.2 - Plaut et. al. model (1994)

The connectionist model devised by Plaut et. al. (1994) is a revised version of the Seidenberg and McClelland (1989) model. By its use of different input and output representations it accounts better for nonword processing.

					Phonology[a]					
onset	s S C		z Z j f v T D p b t d k g m n h			l r w y				
vowel	a e I o u @ ^ A E I O U W Y									
coda	r	l	m n N	b g d	ps ks ts	sz	f v p k	t	S Z T D C j	
					orthography					
onset	Y S P T K Q C B D G F V J Z M N R W H CH GH GN PH PS RH SH TH TS WH									
vowel	E I O U A Y AI AU AW AY EA EE EI EU EW EY IE OA OE OI OO OU OW OY UE UI UY									
coda	H R L M N B D G C X F V J S Z P T K Q BB CH CK DD DG FF GG GH GN KS LL NG NN PH PP PS RR SH SL SS TCH TH TS TT ZZ U E ES ED									

[a]/a/ in POT, /@/ in CAT, /e/ in BED, /i/ n HIT, /o/ in DOG, /u/ in GOOD, /A/ in MAKE, /E/ in KEEP, /I/ in BIKE, /O/ in HOPE, /U/ in BOOT, /W/ in NOW, /Y/ in BOY, /^/ in CUP, /N/ in RING, /S/ in SHE, /C/ in CHIN /Z/ in BEIGE, /T/ in THIN, /D/ in THIS. All other phonemes are represented in the conventional way (e.g., /b/ in BAT). The groupings indicate sets of mutually exclusive phonemes.

Table (4.1) - The phonological and orthographic representations used in Plaut et.al. network. (Plaut et.al., 1995)

Plaut et. al. (1994) argue that what prevented the Seidenberg and McClelland network from exploiting the structure of the English spelling-to-sound system as fully as humans readers, were the limitations imposed by the use of wickelphones for the encoding of graphemes and phonemes. They developed a connectionist model of word naming with a new type of representation in which the generalisation across its training data is structured in terms of formal linguistic categories. Specifically, the input and output levels are divided into onset, nucleus and coda. This means, for instance, that the behaviour associated with the "l" in "silt" does not misleadingly generalise to instances of "l" in onset position, as in "lift". See Table (4.1) below for the phonological and orthographic representations used by them.

An attractor network was used in the simulations. Its architecture consists of three layers of units, as shown in Fig. (4.5) below. The input layer of the network contains 105 graphemes, one for each grapheme in Table I. Similarly, the output layer of the network contains 61 phoneme units. Between these two layers is an intermediate layer of 100 hidden units. Each input unit is connected to each hidden unit which, in turn, is connected to each phoneme unit. In addition, each phoneme unit is connected to each other phoneme unit (including itself), and each phoneme unit sends a connection back to each hidden unit. The weights on the two connections between a pair of units (e.g., a hidden unit and a phoneme unit) are trained separately and need not have identical values. Including the biases of the hidden and phoneme units, the network has a total of 26 582 connections. The states of the units in the network change smoothly over time in response to influences from other units.

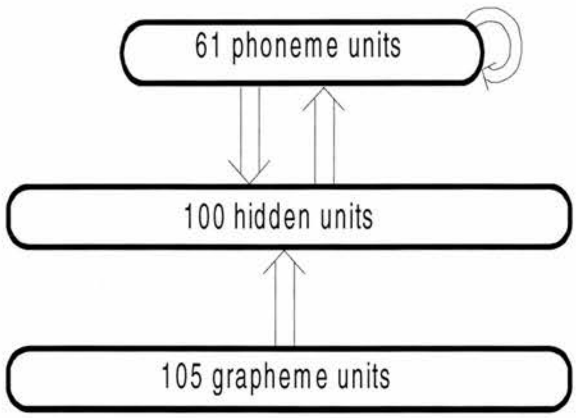

Fig. (4.5) - The architecture of the attractor network. (Plaut et. al.,1995).

The training corpus consisted of the 2897 monosyllabic words used in the Seidenberg and McClelland corpus, plus 101 monosyllabic words missing from that corpus. After training, the network produced 93% of Glushko's regular nonwords, 62.8% of the exception nonwords and 86.3% of McCann and

Besner's control nonwords. However, if any pronunciation is accepted as correct that is consistent with that of a word in the training corpus with the same body (ignoring inflected words and those with final /j/), the network pronounces correctly 97.7% of the exception nonwords, and 92.1% of the control nonwords. Thus, overall, the ability of the attractor network to pronounce nonwords is comparable to that of skilled readers.

4.4 - Motivation for experiments

As we have seen, the category of proper names is viewed as holding a special status in comparison with other category of words. First, experimental and observational data reveal how difficult it is to recall proper names. Second, the vulnerability to ageing effects of our ability to recall proper names. Third, but perhaps the most compelling of all the lines of evidences, the category-specific type of impairment exhibited by patients who have suffered brain injury. All together, these led to the conclusion that proper names are processed somewhat differently from the other types of word. In terms of a model, Valentine et. al (1993), for example, propose "name recognition units" for the processing of names separated from "word recognition units" where all the rest of the lexicon is to be processed.

So far, only those working in the area of face recognition have adjusted their models so as to fit these findings. The more general models of pronunciation however, such as the connectionist models of reading aloud, treat all categories of words equally and ignore any processing differences that might exist between them.

In this chapter we investigate one of the most conspicuous and consistent visual clues in English that belongs to the category of proper names, i.e., initial capitalisation. We hypothesise that the initial capitalisation of a written string, when seen in isolation, i.e., out of the context of a sentence, prompts the reader to process it as belonging to the category of proper names. Brennen hypothesis in his SSPP theory (see page 65), is that the set of plausible phonologies belonging to proper names is smaller than that belonging to nonwords. Here we will be testing Brennen's hypothesis by making use of the experimental paradigm of counting the number of different pronunciations generated by subjects when reading nonwords aloud. In accordance with the SSPP theory we predict that fewer different pronunciations will be generated for the initially capitalised strings by contrast with the non-capitalised ones. Also, from Brennen's theory, we predict that the reaction time for initially capitalised strings will be faster than for the non-capitalised ones. The consequences of confirming this hypothesis are two-fold: first, it would add support, this time in a broader context than that of face recognition and neuropsychological studies, to the view that the category of proper names hold some special status. Second, it also has implications for the present models of reading aloud, in that it would require the incorporation of such features into these models so as to make them more efficient and more psychologically plausible. In the experiment described below we test this hypothesis.

4.5 - Experiments

4.5.1 - Experiment 1

4.5.1.1 - Participants

The participants were 16 volunteers from the Cognitive Science Department of Edinburgh University. All participants had normal or corrected-to-normal vision and were all native English speakers.

4.5.1.2 - Stimuli and design

The stimuli required for this experiment were nonwords, that potentially could support more than one pronunciation. The stimuli were nonwords with a basic monosyllabic structure - onset, nucleus and coda - in which the nucleus was pronounceable in different ways (e.g., the nucleus *ei* of the nonword *seid* has at least the two following pronunciations: /ei/ and /ai/.) An algorithm was developed to construct the stimuli, using onsets, nuclei and codas extracted from the CELEX database, that would reflect all the range of monosyllabic constructions possible.

First, all the nuclei which supported different pronunciations were identified (e.g. "-i-", in "pint" and "mint"; "-ea", in "pear" and "real"). These numbered 27 (21 diphthongs, plus the five vowels and the semi-vowel "y"). The onsets and codas of all monosyllables in the CELEX database were also identified and sorted by their frequency of occurrence. The extremes of high and low frequency onsets and codas were then used together with a selection of the nuclei above to produce 27 nonwords with high-frequency onsets and codas (the *high-frequency* group), and 27 nonwords with low frequency onsets and codas (the *low-frequency* group). Within each group, the onsets, nuclei and codas were randomly combined. Ten other practice nonwords were also created in the same way. The 54 nonwords are listed in the Appendix I.

As it may be seen from Appendix I, this algorithm for creating nonwords produces items that are often orthotactically unusual. Here, we use the expressions *orthotactically regular and orthotactically irregular nonwords*, to refer to those nonwords for which their combined parts respected the orthographic conventions of English and to those that did not, respectively. In the *high-frequency* nonwords, although both onset and coda were high frequency, the random combinations produced by the algorithm included orthotactically irregular nonwords like "coew" alongside the more regular ones like "hean" "jits" or "yauth". In the *low-frequency* nonwords, the low frequency onsets and codas sometimes came from isolated loan words, such as "cz-" from "czar" and may be orthotactically irregular within the onset or coda as well as between onset and nucleus, and between nucleus and coda. There appears to be more potential for different pronunciations of the onset and coda in the *low-frequency* group, where different pronunciations appear principally regarding the nucleus. The *low-frequency* nonwords tended to be longer than the *high-frequency* nonwords and to look more like foreign words than simply unknown or orthotactically irregular English words. Rather than prejudge the issues by excluding any of the nonwords formed by the algorithm, all of the items were employed in the experiment described below, given that such concerns were orthogonal to the principal contrast being investigated, namely the difference between a capitalised and a lower-case form of the same nonword.

Two different conditions, IC(initially capitalised) and NC(non-capitalised) were created. In the IC condition, the first letter of the nonword was upper-case. In the NC condition the nonword was completely written in lower-case. Times font, size 24 points, was used throughout. All 54 nonwords occurred in each of the two sets of materials, with IC/NC conditions counterbalanced between the two sets. Subjects saw each nonword only once. In each set, the *high-frequency* and *low-frequency* nonwords were both divided as equally as possible into IC and NC nonwords. The nonwords were randomised eight times, once for each subject, for the presentation of the materials.

4.5.1.3 – Procedure

The material was presented on a computer screen and the subjects responses were recorded using a digital audio tape recorder and a directional microphone. Instructions appeared on a computer screen at the beginning of the experiment. Participants clicked a mouse to proceed. Subjects were instructed that the experiment dealt with "nonwords: i.e. words that are not normally found in the language". They were instructed to read each word aloud, once only, as quickly, as clearly and as accurately as possible. They were given no other indication as to the nature of the materials.

On each trial, there was a pause of 2000 ms followed by a 350 ms beep, after which the stimulus appeared in the centre of the screen for 2000 ms.

There was then a further pause of 2000 ms before the beginning of the next trial. Participants were allowed to rest halfway through the experiment; the two halves of the experiment were balanced for condition and for nonword type (*high-frequency* and *low-frequency*). Six of the practice items occurred before the start of the main part of the experiment, and two occurred as fillers before each of the two blocks of trials. The practice items were also balanced for condition and nonword type.

Subjects were tested and recorded in a quiet room. The complete experiment lasted less than 30 minutes. The data were transcribed and measured using the Entropic Waves+ program to display the speech files. Approximately half of the total number of subjects, when asked, later reported noticing that some of the words begun with a capital letter.

4.5.1.4 - Results

Two types of dependent variables were used in the experiment: the onset latencies and the number of different number of pronunciations generated by subjects. The two independent variables investigated; the frequency of the components parts of nonwords and initial letter-capitalisation, had two levels each(*high-frequency* x *low-frequency*) and (initially-capitalised x non-capitalised).

A 2(high-low *frequency*) x 2(initial letter-case) ANOVA was performed for the different number of pronunciations[1]. Only F_2 analysis is reported here, since each subject was allowed only one pronunciation per item. A main effect was found for the initial letter-case analysis $F_2(1,52 = 4.06, p < .05)$, such that IC's elicited fewer pronunciations than the NCs. These results support, our prediction concerning proper names having fewer pronunciations than nonwords. The analysis of nonword-frequency however, did not reach significance $F_2(1,52) = 1.34$ $p = .25$. Finally, no interaction was found

[1] See section 3.6.2 for details of calculating the number of different pronunciations.

between frequency and letter-case F_2 (1.52) = 2.07, p = .16. See Table (4.2) below for the mean number of pronunciations and standard deviations.

	IC Initially capitalised		NC Non-capitalised	
	No. of different pronunciations	SD	No. of different pronunciations	SD
High-frequency	2.59	1.18	3.04	1.53
Low-frequency	3.18	1.27	3.37	1.52

Table (4.2) - Measurements of the number of different pronunciations and their standard deviations. Two variables, nonword-frequency and initial-capitalisation were tested.

A 2(letter-case) x 2(*frequency*) ANOVA was also employed for the onset latencies. This time the analysis was performed with both, subjects (F_1) and items (F_2). A main effect for the frequency variable was found $F_1(1,13) = 34.7$, $p < .001$ and $F_2(1,52) = 10.2$, $p < .003$ in the analysis. It took longer to react to *low-frequency* nonwords compared to the *high-frequency* ones. The letter-case analysis did reach significance for $F_1(1,13) = 6.04$, $p =.03$. However, it is not significant for $F_2(1,52) = 0.08$, $p = .37$. This outcome is discussed in the next section. Finally, no interaction was found between frequency and letter-case either for $F_1(1,13) = 1.75$, $p = .21$. or $F_2(1,52) = 0.01$, $p = .91$. See Table (4.3) below for the means and standard deviations.

	IC Initially capitalised		NC Non capitalised	
	RT(ms)	SD (ms)	RT (ms)	SD(ms)
High- frequency	841.2	173.8	811.5	148.2
Low-frequency	953.9	200.5	930.9	163.3

Table (4.3) - Measurements of the onset latencies and their standard deviations. The same variables in Table (4.2) above were tested.

4.5.1.5 - Discussion

Two different ways of measuring subjects' performance were adopted for this naming task. The more traditional *onset latency* measurements were performed as well as a *count of the number of different pronunciations* produced by participants when naming the strings aloud. The second type of measurement was possible by careful design of the nonwords, which could have more than one pronunciation.

A significant effect was found for the letter-case variable corresponding to the onset latency measurements. In other words, subjects' reaction times have been affected by the fact that half of the stimuli were initially capitalised and the other half were not. Some support for these results can be found in Baron (1977), where, in one of his experiments, he detected a familiarity effect related to the initial capitalisation of proper names. He asked 31 subjects to name a list of thirty names of three letters or less, in a familiar form (Al, Abe, Ann, Dan, Ed, Don, etc.) and in an unfamiliar form (al, abe, etc.) He found a significant effect of familiarity for names.

The failure to obtain significant F_2 results might be due to the type of measurement used here. Reaction time measurements might not be sufficiently sensitive to uniformly capture the delicate effects of initial capitalisation. Also, our policy of not restraining the type of materials used in the experiment might have added too much variability.

However, there is a long-standing debate in the literature of how much perceptual information is used in word recognition (Paap et. al., 1984). Interestingly, studies which use general alternation of case find no such effects: as, for instance, it is reported by McClelland (1976) and Adams (1979). They have shown that in tachistoscopic reports the word-nonword advantage was unaffected by case alternation. More recently, Mayall and Humphreys (1996) also report not finding any "shape" effects in a naming task where alternated letter case strings were used as stimuli.

A further finding concerning the reaction time measurements that deserves mentioning here is that of subjects taking longer to react to the initially capitalised items compared to the non capitalised ones (see Table 4.3). At first this might seem to be in contradiction with the finding that a smaller number of different pronunciations have been generated for the initially capitalised nonwords in comparison to the non capitalised ones. However, this result needs not to be contradictory if it is assumed, as we do here, that the retrieval of proper names phonology is more costly than that of common nouns. The reason for that, we advance is that time is spent in supressing the existing phonology. Support for this hypothesis is arguably to be found in the tip-of-the-tongue phenomenon, where proper names are reported to be notoriously more difficult to retrieve than common nouns.

Another type of variable manipulated in Experiment 1 was frequency: strings composed of only high-frequency parts were *called high-frequency nonwords* and those composed of low-frequency parts were called *low-frequency nonwords*. A significant difference in onset latencies was found between the *high-* and *low-frequency* nonwords. It took longer for subjects to name the low-frequency strings as compared to the *high-frequency* ones. Although the results are in the expected direction, they should be viewed with caution, since the nonwords used in the experiment were not controlled for either onsets or for the number of letters forming the strings. As it can be seen from Appendix I, the *low-frequency* strings are longer than the *high-frequency* ones and also have a more "foreign" appearance. It should be noticed however, that no difference was found for the frequency variable regarding the counting of the different number of pronunciations.

Finally, the results obtained by the counting of the number of different pronunciations generated by participants found that initially capitalised nonwords presented a smaller number of different pronunciations than the non-initially capitalised ones. This result is at the heart of the motivation for Experiment 2, to be described next. We defer a full discussion of it to the general discussion at the end of the chapter.

4.5.2 - Motivation for the second experiment

In Experiment 1 we obtained the intriguing result that subjects produced a smaller number of different pronunciations when presented with initially capitalised strings in comparison to the number of pronunciations produced in the presence of non-capitalised ones. These results were obtained in an experimental situation where subjects' performance was being monitored under time constraints. In the experimental instructions they were solicited to name the stimulus flashing on the screen as quickly as possible. In our next experiment, we look at subjects' performance in a similar task, but this time with no time constraints. In this experiment subjects participate in a self-paced naming task. Subjects were instructed to produce as many different pronunciations as they felt appropriate for the stimulus. Our aim was to replicate the result we found earlier, that is a smaller number of pronunciations for initially capitalised nonwords, now that there were no time constraints.

The equipment used to present the stimuli and to record subjects' responses was the same as for Experiment 1. In Experiment 1, we deliberately avoided pre-theoretical selection of nonwords. This allowed us to explore the effects of initial capitalisation on a extensive range of materials. Recall that the algorithm presented in Chapter 3 was used to sample as widely as possible from the range of possible nonwords. In Experiment 1, stimulus materials were not excluded, for example, because they contained unusual orthographic transitions or because they resembled foreign words. In Experiment 2, however, stimuli were carefully chosen so as to resemble as closely as possible normal words in English. This was done to obviate any claim that the effects obtained in the Experiment 1 were artefacts, i.e., a product of the unusual nature of the material used there. Also, only high-frequency onsets and codas were used to compose the nonwords. This was done in order to avoid nonwords that would exhibit unusual orthographic transitions. We also almost trebled the number of items in the second experiment (from 54 items in Experiment 1 to 142 in the Experiment 2). We also noticed that some of the items used in our first experiment did not strictly fit our description of being "monosyllabic", for example, "chomy" and "byaeck". In the Experiment 2 we amended the algorithm so that at least one monosyllabic pronunciation was possible for every item.

In summary, the aim of the new experiment is to test whether participants, when performing a self-paced naming task with nonwords that resemble the orthography of English more closely, would generate similar results to those obtained in the first experiment with respect to the number of different pronunciations produced. If so, it would be adding support to the hypothesis that proper names really hold a special status insofar as their processing is concerned.

4.5.3 - Experiment 2

4.5.3.1 - Participants

The participants were 26 students from the Linguistics Department of Edinburgh University. All participants had normal or corrected-to-normal vision and were all native speakers of English.

4.5.3.2 - Stimuli and design

The stimuli used here were built with the help of the same algorithm used in the first experiment. All the chosen stimuli were nonwords that potentially could support more than one pronunciation. These are listed in Appendix II. All items were chosen to be either monosyllabic or have at least one monosyllabic pronunciation. All orthotactically unusual items, i.e., all those items that do not seem to conform to the orthographic rules of English, were disregarded, so that all nonwords resembled as closely as possible normal words in English. The total number of nonwords constructed for the experiment was 144. The nonwords were composed of either five or four letters. From this total 72 of them are "weird" and 72 are "nonweird". The reader is referred to Chapter 2, for discussion on the concept of *weirdness* employed here.

The vowels used to compose the nucleus were only those that could be read by Plaut's neural net, according to Table (4.1), in section 4.4.3.2..Two different conditions, initially capitalised (IC) and non-capitalised (NC) were created. The 144 nonwords were grouped into two sets of materials, with IC/NC conditions counterbalanced between the two sets. In each set, the weird and nonweird nonwords were both divided as equally as possible into IC and NC nonwords. Thirteen different random orders, one for each subject were generated for the presentation of the materials. The type of vowels used in each nonword was controlled for each participant. If a participant saw a capitalised nonword that contained the vowel “au”, he also saw a different nonword that also contained the vowel “au”, but the latter would not be capitalised.

4.5.3.3 - Procedure

Participants were instructed that the experiment dealt with “nonwords”, i.e., “words that are not normally found in the language”. They were instructed to read each nonword aloud and to give as many different pronunciations as they felt possible. They were also told to do it as clearly and as accurately as possible. They were given no other indication as to the nature of the materials.

On each trial, there was a pause of 2000ms followed by a 350 ms beep, after which the stimulus appeared in the centre of the screen for as long as the participant felt it necessary. The next trial began with the subject clicking the mouse. Participants were allowed to rest halfway through the experiment. The two halves of the experiment were balanced for condition and for nonword type (weird x nonweird). Six of the practice items occurred before the start of the main part of the experiment, and two occurred as fillers before each of the two blocks of trials; practice items were also balanced for condition and nonword type.

Participants were tested in a quiet room. Instructions appeared on the screen at the beginning of the experiment. Subjects clicked the mouse to proceed. The complete experiment lasted about 45 minutes.

4.5.3.4 - Results

A 2(initial letter-case) x 2(weirdness) ANOVA was used to determine any statistically significant differences in the number of pronunciations. The F_1 analysis was not statistically significant for either of the variables, i.e., weirdness $F_1(1,25)= 0.20$, $p = .66$. and capitalisation $F_1(1,25)= 2.16$, $p = .15$. However, a main effect of weirdness was found for $F_2(1,140) = 12.9$, $p < .000$). A larger number of pronunciations were found for the weird condition. A main effect concerning the letter-case variable was also found $F_2(1,140) = 3.91$, $p = .05$. A smaller number of different pronunciations was obtained for the initially capitalised nonwords. Finally, no interaction was found between weirdness and initial letter-case $F_1(1,25) = 0.25$, $p =.62$ and $F_2(1,140) = 0.01$, $p = .92$. See Tables (4.4) and (4.5) below for the means and standard deviations.

	IC Initially capitalised		NC Non capitalised	
	no. of different pronunciations	SD	no. of different pronunciations	SD
Weird	6.75	3.28	7.03	3.34
Non-weird	4.90	3.02	5.15	3.12

Table (4.4) - Measurements of the number of different pronunciations and their standard deviations. Two variables, nonword weirdness and initial capitalisation were tested (Item analysis

	IC Initially capitalised		NC Non capitalised	
	no. of different pronunciations	SD	no. of different pronunciations	SD
Weird	44.8	14.3	43.8	12.1
Non-weird	42.5	11.5	45.5	15.8

Table (4.5)- Measurements of the number of different pronunciations and their standard deviations. Two variables, nonword weirdness and initial capitalisation were tested (Subject analysis).

4.5.3.5 - Discussion

A peculiarity of the results above is the contrast in significance between the results concerning subject and item analyses. The subject analysis for both variables did not yield significant results. Although all subjects were native speakers of English and an effort was made to have as homogenous a population as possible, it could be the case that a number of them were not completely monolingual. We speculate that this fact could have interfered with their performance obliterating the effects concerning the two variables.

Both variables are statistically significant for the item analysis. Regarding capitalisation, the results confirm those of the previous experiment: when nonwords are seen with initial capitalisation, they are pronounced with a smaller range of different pronunciations. For the second variable, a significant difference was found between the *weird* and the *nonweird* nonwords. The weird nonwords have a larger range of pronunciations than the nonweird. These issues will be discussed at length together with the results of Experiment 1 in the general discussion.

4.5.4 - General discussion

The view that there exists dissociable special-purpose systems for handling the syntactic aspects of language on the one hand and the semantic content on the other, is well documented in the neuropsychological literature (Caramazza & Berndt, 1978). Furthermore it is not unusual to find claims that some categories of words have different storage and processing: recall that the most prominent models of face recognition have adopted the view that proper names are a special category of words and as such are processed separately from the main lexicon. Here we present further evidence of special processing strategies specific to particular categories of words.

An abundant amount of research has also been generated, for example, concerning differences between open (content) versus closed (function) class words. Bradley and Garrett (1983) proposed the existence of separable specialised recognition devices, one for dealing with open class items and one for closed classed items. Shillcock and Bard (1993) argued that syntactic processing intervenes in lexical access and predetermines the potential competition between the function and content words homophones (e.g., *would/wood*). Also in a more recent work, Shillcock and Kelly (1997), have shown by using the clarity gating technique, that function words by contrast with content words are the only ones to present significant effects of lexical neighbourhood for the identification of high-frequency words: larger neighbourhoods inhibited recognition. They claim this finding can be predicted by a model that allows strong interaction between syntactic processing and lexical representation of function words, and in which identification of words relies more on context and less on sensory input compared to content word identification.

The idea that contextual information - specifically, syntactic information - can influence the reading of isolated nonwords is also suggested by Besner et. al. (1990) in the discussion of the Seidenberg and McClelland model and in a reference to work by Campbell and Besner (1981). Campbell and Besner showed that the pronunciation of nonwords that began with the letters "th-" was strongly influenced by their perceived syntactic context. If a nonword like "thuz" occurs in a closed-class syntactic position, as in (1) below, it is pronounced with a voiced initial fricative, whereas if the context is that of an open-class word, as in (2), the nonword is pronounced with an unvoiced initial fricative.

(1) I can see four cans over thuz near a tree.
(2) Mary's thek is very nasty.

This behaviour reflects the fact that an initial-th-closed-class word in English is pronounced with an initial /ð/, whereas an initial-th-open-class word is pronounced with an initial /θ/. This syntactic effect on pronunciation is robust; indeed, it even survives in nonfluent dysphasia (Shillcock & Hacket, in press).

We return now to proper names and the view that they are part of a category of words that holds some special status concerning their storage and processing. We decided to explore the fact that in languages such as English, the vast majority of proper names are represented using initially capitalised letters. One of the motivations for running the present experiments was to verify if the use of initial capitalisation in isolated nonwords, would cue readers to behave as if perceiving them as belonging to the category of proper names.

In both experiments, subjects generated a smaller number of different pronunciations for the initially capitalised strings (IC) in comparison to the non-capitalised ones (NC). The *set size of plausible phonologies* (SSPP) theory mentioned in section 4.2.2.3, neatly explains these findings, if it is assumed, as it is in this thesis, that initial capitalisation is used by readers as a clue to the processing of proper-names. As we have seen, SSPP was put forward by Brennen (1993), to try to explain why proper names are more difficult to recall than common words. He argues that the set size of plausible phonologies for proper names is larger than that for common names. Note also that, the set size of plausible phonologies for nonwords was hypothesised to be much larger than that of proper names. In summary, SSPP (nonwords) >> SSPP (proper names) >> SSPP (common English words). In fact, when discussing some experiments in the light of the SSPP theory, Brennen asserts that he would expect people's names to be better recalled than nonwords.

We report here on nonwords; half initially-capitalised nonwords and half non-capitalised nonwords. According to our predictions, subjects should interpret the initially-capitalised nonwords as belonging to the category of proper-names as opposed to the non-capitalised ones. If this assumption is correct then one would expect a smaller set size of plausible phonologies to be available for the initially capitalised group in contrast with the other group. Thus,

$$\boxed{\text{SSPP (IC)} = \text{SSPP(proper names)} << \text{SSPP (nonwords)} = \text{SSPP(NC)}}$$

Therefore, it should be concluded that a smaller range of different pronunciations would be generated for initially capitalised nonwords. This is exactly what was found experimentally both in the weirdness x initial capitalisation and in the frequency x initial capitalisation experiments. Thus, we argue that the present findings are a corroboration of the psychological reality of the SSPP theory. The two experiments above demonstrate that simply changing the initial capitalisation of a nonword is enough to produce significant changes in the way it is processed by the brain.

These findings corroborate our main hypotheses that names are processed differently.

Let us now examine the consequence of these findings for the models of reading aloud, more specifically, for the connectionist models.

Clearly, the two most influential distributed models of word recognition, i.e., Seidenberg and McClelland (1989) and Plaut et. al. (1993), have no means of taking syntactic constraints like those referred to above into account. So, information of the type that Besner and Campbell (1981) talk about, is overlooked by the models and they generate /θ/ and /ð/ from "th-" and its lexical context using the same criteria as for any other orthography to phonology mapping. Besner's observation regarding syntactic context stands as an example of how pronunciation is influenced by formal linguistic distinctions not wholly present in the orthography of the word being pronounced.

Now, in the case of the initial capitalisation of proper names, models of pronunciation with distributed representations represent viable models for discussing the possible relationships of proper names to the rest of the words in the lexicon. They are able to simulate the basic data from studies of the effects of lexical neighbourhoods, in which the processing of low frequency words benefits from the existence of similar words in the lexicon (Andrews, 1989, 1992; Voice, 1995). Proper names do not feature in the training sets of such published connectionist models, but the simplest means of incorporating them is as ordinary words, with no distinction made between upper- and lower-case letters. For example, "Smith" and "smith" would have only one input representation and "Fred" would receive an input representation in which the grapheme standing for "F" also stands for "f" in "flat". This proposal would mean that the lexical neighbourhood for "free" would include "Fred" as well as "flee", "tree" and "fret", and that capitalising a nonword should have no effect on pronunciation, as the same conspiracy of neighbours would determine processing. However, according to what has been discussed here and shown by our experiments, the idea of a single space shared by the whole lexicon that equally serves all the different categories of words, does not seem to apply. Therefore, we propose that distributed models should incorporate into their structures, means of restraining their process of generalisation into a restricted lexical space, where the grouped items are all exemplars bearing phonologics that are plausible to that domain. Furthermore, they should find a way of incorporating mechanisms that deal with specific aspects of specific categories. In the case of proper names, the processing of visual cues are an integral part of manipulating proper names.

Earlier two different approaches to language modelling have been briefly discussed, i.e., the box-and-arrow and the biologically based types of model. Although the view taken here is that the biologically based models are our best hope for an explanation of such complex phenomena as those happening in the brain, we have also fully acknowledged the important role that is played by the box-and-arrow models. Perhaps, their contribution to cognitive psychology is that of making very clear the multitude of issues involved in modelling, in their postulation of the many different "levels" of processing. A third type of modelling, that could be seen perhaps as intermediate between the two mentioned above, is the connectionist approach. On the one hand, connectionist descriptions are of more abstract levels than the biologically based models, on the other hand, the type of mechanism they propose aims to resemble brain architecture at least stylistically. Unfortunately, the biologically based models are still very much in their infancy to be used here as a means of explanation for such subtle processing strategies as that of the influence of capitalisation in visual word recognition. However, connectionist models are rapidly evolving and adapting to respond to the new circumstances in cognitive psychology, especially those dealing with language architecture. Our attempt to offer a more concrete mechanistic explanation of the subtleties of initial capitalisation cueing in reading, has as its main source of inspiration the latest model of reading by Harm & Seidenberg (in press).

This model can be considered as a further development of two earlier models of reading aloud, i.e., the Seidenberg and McClelland (1989) and the Plaut, McClelland, Seidenberg and Patterson (1996). See sections 4.3.3.1 and 4.3.3.2 respectively for more details on these models. The main goal of its development was the investigation of the role of phonological information in early reading and dyslexia. Phonological information plays a central role in learning to read and in skilled reading (Adams, 1990). Children's knowledge of the phonological structure of language is a good predictor of early reading ability (Bradley & Bryant, 1983; Shankweiler & Liberman, 1989). The model was exposed to phonological word forms and learned to represent them in memory, so as to equate to a child's acquisition of phonological knowledge prior to learning to read. The manner in which the network represented this information allowed it to extract generalisations about the phonological structure of English; in particular, it learned about the structure of phonemic segments and about constraints on the sequences of phonemes (phonotatics). The phonological representation scheme that was employed has two principal design features. First, it employs a distributed representation of phonemes in which units correspond to phonetic features. Second, each phonetic feature unit was connected to every other and to a set of phonological "cleanup units".

After training, the model's phonological representation was taken to approximate a beginning reader's knowledge of phonology and the next step was the learning of the mappings between orthography and phonology. In order to do that, to the trained set composed of cleanup units and phonological units was added an input layer composed of 208 units representing the spelling of words. These were fully connected to an intermediate level of 100 hidden units, which in turn were fully connected to the output representation, which was the phonological attractor net mentioned above. The model was trained to map the spelling of a word onto its pronunciation both with and without the pre-trained connections to the cleanup units. See Fig. (4.6) for a schematic view of the net's architecture. The fundamental point made by this model is that phonological pre-training facilitates the mapping between orthography and phonology.

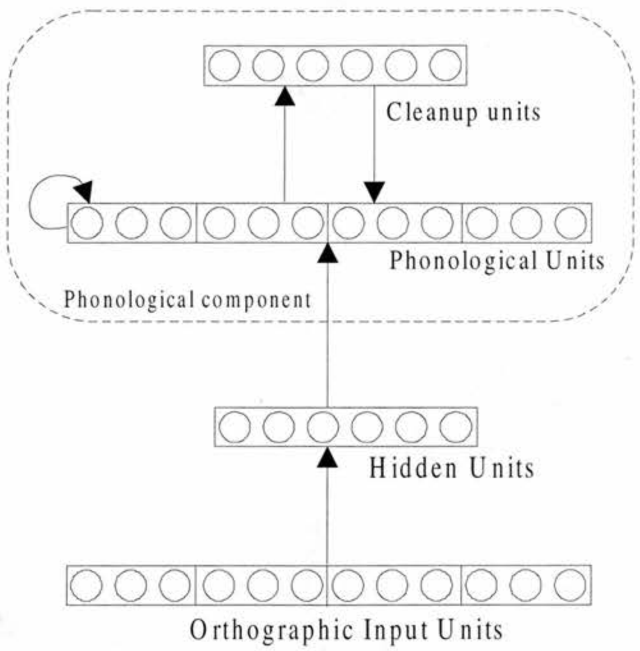

Fig. (4.6) - The architecture of Harm & Seidenberg reading model. (in press)

As in Harm & Seidenberg's model, an essential feature of the type of mechanism we propose here for explaining the effects of initial capitalisation in reading is the so-called "cleanup units" device used by the phonological attractor network. Let us suppose, that by using the simple device of turning a capitalisation unit "on or off" the net is able to make distinctions between a string that is represented by a capital initial letter and a string in which all letters are represented in lower case. Now, most of the time the rich variety obtained for pronunciation of English strings is caused by the wide scope with which vowels can be pronounced in that system. Let us suppose further that the shape of the competition in the net at the phonological level is regulated by the type of orthographic input i.e, if the string is initially capitalised or not. The resulting competition will be more varied if the input representation has an initial capital letter and in this way allows for a more adventurous pronunciation of vowels. As an illustration, if the input string is a real word such as "bucket" and it has a complete lower-case representation, the choice of phonological units associated with it is strongly constrained by its lexical status as a common English word and also by its relatively high frequency. As a result, the connections between its orthographic representation and the correct phonological mapping have weights that will make it easy for them to win any competition posed by other phonological units. The net rapidly enters into a stable state.

On the other hand if the string "Bucket" has an initial capital letter fewer constraints are placed on its choice of pronunciation, since there are no strict phonological rules that apply to names, in other words, the set of alternative phonologies for proper names is larger than that of common words, thus allowing for more scope in the choice of pronunciation of a string. Names, as we have discussed earlier, are of a special nature and compared to common words they are associated with massive amounts of pragmatic, semantic and emotional connotations. Therefore, a purely orthographic- phonological mapping of name pronunciation can not do justice to their processing and it is bound to fail in simulating accurately human performance. We therefore propose that there is also semantic/contextual input to the phonological units, as has always being envisaged by Seidenberg et. al. (1989), to deal with homophones such as *wind/wind*, to distinguish sentence-initial capitalisation from proper names, and perhaps to mediate generalisation about name pronunciation such as that discussed by Lindsey(1990). He discusses the process by which speakers assign pronunciations to names and loan words and observes that for RP English the strategy seems to be one of assigning short/lax vowels by default to closed syllables, as in newly encountered names or loan words (e.g. "lasagne", "macho", "kebab"), but with an option for lengthening the vowel, as in "Mahler", "Evita", "diva". (In contrast, in Modern General American pronunciation, the strategy seems to be one of assigning a long/tense vowel - "Nicaragua", "Nissan".)

A question now is how the net behaves if a simulation is run with the same type of materials used in the experiments reported earlier in this chapter? The two types of materials were nonwords and fictitious proper names, i.e., nonwords that start with an initial capital letter. For the fictitious "proper names" a smaller range of different pronunciations will be assembled for competition in comparison to those assembled for the nonwords, since the domain of nonwords is less constrained phonologically speaking. The time taken to output the initially capitalised nonwords is also going to be longer, since the processing of "alternative phonologies" is more costly due to the very low frequency of the phonemes involved and also by the fact that the more familiar patterns of segmentation and phonotatics of English might be in conflict with those required to pronounce the alternative phonologies.

In the rest of our discussion we turn to the results that were obtained through the manipulation of the variable called *weirdness*[2]. The concept of weirdness has been dealt with in Chapter 2 and is defined in terms of neighbourhood density. Weird nonwords are by definition neighbourless, since the combinations of their composing parts are **not** present in the lexicon. The nonweird nonwords are composed of parts that **are** present in the lexicon and thus can not be neighbourless. In Experiment 2, we have divided the stimuli so that half are weird nonwords and the other half are nonweird.

It was also found that the capitalisation and weirdness variables did not interact. A count of the number of different pronunciations shows that weird nonwords always produced a larger number of different pronunciations than the nonweird. As an example, we quote the non-weird nonword "fier", which has the following orthographic neighbours: "bier", "pier" and "tier". A total of four different pronunciations were generated by the subjects for the nonword "fier". The weird nonword "hoess" had a total of ten different pronunciations. These results can be interpreted as a direct consequence of lexical competition. On the one hand, the pronunciation of the non-weird nonwords is constrained by the pool of neighbours surrounding them. On the other hand, the poverty of the lexical space inhabited by the weird words frees them to be pronounced more adventurously. The effects of neighbourhood density is a popular topic in the psycholinguistic literature. Coltheart et. al. (1977), for example, has shown that nonwords that have many neighbours take longer to classify in a lexical decision task, than nonwords that have none or just few neighbours. Andrews (1989) found for naming and lexical decision tasks, that facilitatory effects of neighbourhood size were observed for low but not high frequency words in all tasks except delayed naming. Voice (1995) showed that lexical decision for isolated words is slowed when a target word has either many orthographic competitors, or a single strong competitor.

[2] The motivation for using weirdness as a variable in Experiment 2 come from the behaviour that was observed for a small part of the materials used in Experiment 1 as a result of a ad-hoc analysis.

The concept of neighbourhood has also emerged in the context of orthographic priming, that is priming due to similarity of orthographic form. It has been established that when two words are presented in rapid succession the processing of the target is facilitated by the prime if they are closely related orthographically, i.e., if they are neighbours (Evett & Humphreys, 1981; Forster & Davies, 1984). Even more relevant in the present context is the suggestion put forward by Foster and Taft (1994) that neighbourhood density should be defined in terms of both individual letter units and subsyllabic units and that both types of density jointly determine priming.

We finish this discussion by raising the issue of initial capitalisation in English not being solely the privilege of proper names. Common words are normally initially capitalised to signal that they are beginning a sentence. We investigate this issue in chapter 5 and make some suggestions as to its relevance.

4.6 - Conclusions

In conclusion, we have seen from many of the sources of evidence available (experimental and behavioural studies, neuropsychological cases etc.) that the organisation of the mental lexicon seems to be very complex. One of these complexities seems to be the special status held by some category of words. The category of proper names seems to be one of them. As a rule, initial capitalisation is used in written English to represent proper names. In our experiments, we have shown that initial capitalisation is a clue strong enough to prompt subjects to treat initially capitalised nonwords differently from others which are not capitalised. It was shown that the subjects produced a smaller number of pronunciations for the initially capitalised nonwords than for the non-capitalised.

We interpreted these results in the light of Brennen's set size plausible phonologies account and have given our interpretation of the results in terms of a connectionist model of reading where the initial capitalisation plays a role in the net's phonological output. We have also investigated previous findings that neighbourhood density is a determining factor in pronunciation by using nonwords manipulated according to frequency and weirdness. We obtained similar results as those found in the literature. As before, the number of different pronunciations generated by subjects was also measured in this experiment. We found that larger numbers of different pronunciations were produced for nonwords with more sparse neighbourhoods (weird group) than for the nonweird group.

Chapter 5

Visual mechanisms implicated in familiarity assessment

In the previous chapter we have offered some evidence that initial capitalisation is implicated in the processing of visual word recognition. We have hypothesised that because in English proper names are most of the time represented with an initial upper-case letter, this fact acts as a clue to the category of proper names. In the present chapter we continue our investigation by looking at other familiar and unfamiliar case representation in English.

Normal readers are extremely skilful at switching between the different fonts or cases they encounter in their reading environment. The issue of how much these different types of visual information are involved in the actual task of reading has generated a variety of theoretical controversies. In this chapter I first consider the types of visual information that might be used to identify words. Then, I briefly describe Besner and Johnston's (1989) model of visual word recognition, focusing on the visual familiarity assessment and the orthographic familiarity assessment routes. I report an experiment with nonwords that was carried out using the same-different matching task paradigm. Next, I explore the idea of using the orthographic familiarity route of Besner and Johnston (1989) to explain our experimental results. I then propose some of the possible mechanisms that might be in place allowing for the use of the route mentioned above. In the present chapter I also discuss the experimental results concerning the two different types of nonwords that were used as stimuli: the weird and the nonweird nonwords. Finally, in a postscript to the chapter I give a reinterpretation of a neuropsychological case involving a patient with a deficit in initial letter identification.

5.1 - Two views: the holistic and the analytic

The question as to what is the code that mediates the recognition of printed words has been around for some time now. It can be traced as far back as the last decades of the nineteenth century (Cattell, 1886). However, so far, no consensus has been reached on the subject. One of its central tenets is the dispute between the analytic and holistic views. The analytics argue that word recognition relies on preliminary letter identification alone (not to be confused with first letter identification). They deny the role of any other feature during the extraction process (for example, word-shape), asserting that the only evidence used in word recognition is the ordered identities of the component letters (Adams, 1979; Henderson, 1982; McClelland, 1976). Some go further adding that it is the **abstract** identity of the individual letters that matters in word recognition, (Allport, 1979; Coltheart, 1981; Evett & Humphreys, 1981; Besner et. al. 1984; Besner & McCann, 1987; Paap, Newsome & Noel, 1984). The holists, on the other hand, advocate that the identification of individual letters is not the only type of information that plays a role in word identification. There are also, according to them, other types of visual information that might be involved, e.g., the outline (or envelope) of the word (Cattell 1886; Crowder, 1982; Monk & Hulme, 1983; Haber & Haber, 1981; Haber, Haber & Furlin, 1983) or the entire set of visual features in words (configurational features) (Wheeler, 1970; Rumelhart & Siple, 1974; Coltheart & Freeman, 1974; Rudnick & Kolers, 1984; Masson, 1986; Howard, 1987; Smith 1969; 1971; Henderson, 1982).

5.1.1 - The holistic approach

The belief that the whole word is the functional unit of word recognition, finds its beginning in the works of Cattell (1886) and Erdman and Dodge (1898). Cattell found that some words can be named more quickly and more accurately than single letters. Erdman and Dodge (1898) reported that the exposure duration necessary to identify four to five letters in a display was sufficient to read single words that could contain as many as 22 letters. These results led them to the conclusion that words are processed as a unit. However, there are alternative interpretations of their data. One is that the context in which a letter is presented influences not the process of perception itself, but only the accuracy of post-perceptual processes. For example, subjects might have been more successful with words because they used sophisticated guessing strategies during the experimental task, i.e., they might have made use of any available information to facilitate letter processing. Such information is unavailable when single letters are presented. Another possibility, is that due to the lack of proper masking during experiments, the word superiority effect could be attributed to failures in short-term memory.

Although both works have many flaws, they deserve credit for having inspired a profusion of relevant research in the area. These are based on many different techniques and paradigms and bear an intricate pattern of results. I refer to Henderson (1982) and Paap et. al. (1984) for a thorough early review on the literature of the subject.

5.1.1.1 - The word-envelope

The term envelope is used to refer to cues from the overall contour of a word. It is variously known in the literature, as word-outline, word-shape and also as low spatial frequency information. As an example of word-envelope, the word *pool*, starts with a descender and ends with an ascending letter. Thus, its envelope would be loosely represented by something similar to the picture below.

Fig.(5.1) - The word pool and its envelope

However, the validity of evidence for the use of envelope cues is questionable since it is open to alternative explanations. The issues of confusability, longer processing time and the lack of advantage related to unique shapes are the most criticised problems surrounding the experimental evidence for word-envelopes.

Confusability refers to the fact that envelope similarity is highly correlated with individual letter similarity, posing a problem for the existence of independent word envelope effects. The finding that subjects perform better at proof-reading words which have had their shape altered by the misspelled letter (e.g., tesc for test) when compared to those which have had their word shape preserved by the misspelling (e.g., tesf for test) has frequently been cited as evidence for word-envelope information playing a role in visual word recognition (Haber & Schindler, 1981; Monk & Hulme, 1983). An alternative explored would be to assert that instead the locus of the effect resides at the level of individual-letter identification. Ponansky and Rayner (1977), for example, demonstrated that picture naming was facilitated both by letter sharing and envelope sharing nonword primes. Also, Paap et. al. (1984), ran an experiment where they varied both letter confusability and word shape and found that all the apparent effects of word shape were due to the confounding factor of letter confusability.

In normal circumstances, the time taken to recognise a word is very short. It is generally acknowledged that word recognition is almost automatic, taking around 200 ms to be accomplished (Rayner, 1989). Haber, Haber and Furlin (1983) ran an experiment where they found that subjects' performance improved 20% when they were cued for word shape and length, in dealing with a task of guessing words in sentences that ended up randomly on a computer screen. This finding was seen as evidence for word-shape information playing a role in word recognition. However, as Besner and

Johnston (1989) remind us, the validity of this conclusion is questionable, due to the unlimited amount of time that was allowed to subjects for generating the guesses. It is possible that they were using different strategies from those normally used during the almost automatic process of visual word recognition.

If word shape has any importance in word recognition, reasoned Paap et. al. (1984), then the number of words that share the same shape must be an indicator of that. The smaller the set of words that share a type of envelope the faster and more accurate word perception can occur with only a little additional information about individual letters. On the other hand, in cases where the shape matches that of many different words, the recognition has to rely more on letter identification than on shape. Therefore, words with relatively rare shapes should generate a better performance compared to those with relatively common shapes. Paap et. al., (1984) set out to verify the issue by running two experiments, a lexical-decision and a tachistoscopic report with words that have rare shapes (e.g. kept, health and death) and words that share very common shapes (e.g. dead, deal, loaf, leak...). They did not find any difference in performance between rare shapes and common shapes and concluded that word shape is not relevant to word recognition.

5.1.1.2 - Configurational features

There is no precise definition in the literature as to what configurational features are. Besner and Johnston (1989) for example refer to them in the following way:

> "The idea that word identification is mediated by what we are calling a word-specific visual pattern is actually quite simple. It amounts to the claim that a word can be identified in the same way as other simpler forms - that is, on the basis of its component visual features. These features might simply be the set of all component individual strokes in the proper arrangement (Smith, 1969), or they might be more exotic aspects, such as junctions between strokes, the shape of spaces between letters (Wheeler, 1970), or any other property of the pattern."

In the same manner as with word-envelope, configurational features are also designated in the literature by a variety of other names such as transgraphemic features, spatial frequency features, or as above by WSVP (word-specific visual pattern).

Possibly, the most elegant way of describing them is as spatial frequency features. In a string of letters, the highest spatial frequency features are the vertical, horizontal and sloping lines. Of a lower frequency, are perhaps, the individual letters and spaces between them. Going one step further, one can say that graphemes containing more than one individual letter have a still lower frequency. It is possible to imagine that groups of graphemes can repeat in certain words, which would constitute a very low spatial frequency configurational feature. The lowest possible configurational feature in this approach would be the whole word envelope. Take for example the word *maintain*. It is clear that the vertical lines are the most numerous. Then there are the individual letters and spaces. The grapheme *ai* has a still lower spatial frequency. In this word we don't have repeating groups of graphemes, which would be the next level of low frequency features. The lowest spatial frequency feature for a word is its envelope.

One source of supporting evidence for the role of configurational features in word recognition comes from the field of neuropsychology. Howard (1987), describes a patient , T.M., with an acquired dyslexia that was a consequence of a cerebral vascular accident. Besner and Johnston (1989) summarise the important facts about Howard's patient in the following terms:

> "T.M. cannot orally read nonwords, can name virtually no letters (1/20) and is extremely poor at pointing to a letter from four visually displayed alternatives when the target letter is presented auditorily (7/20). He is also very poor at cross-case matching of single letters (e.g. A/a-yes; A/r-no).
>
> In contrast, T.M.'s oral reading of single words, while far from perfect, seems to exceed what might be expected if word identification depended only upon preliminary letter recognition (32 percent of a set 1002 words were correctly read, and a further subset of errors were semantically related to the target). Further experiments found that inserting plus signs between letters (e.g. p + l + u + s)

> drastically impaired T.M.'s oral reading of single words (4 percent correct vs. The 32 percent with normal presentation). Case alternation also impaired performance, although less severely (14 percent correct). T.M. was also completely unable to correctly read abbreviations. His error responses were, however, semantically related to the target on 14 of 30 trials when the stimulus was displayed in familiar visual format (FBI; RSPCA), versus 1 of 30 trials when the stimulus was visually unfamiliar (fbi; rspca). Howard concludes that T.M.'s letter-identification abilities are very poor indeed."

These data are impressive and worth mentioning, because they show that the patient above seems to be relying upon the shape of the words, to be able to read them aloud. However, it is problematic to generalise these results to the type of processing used by people with normal reading abilities. The most that can be said here, is that these types of strategies are accessible to the brain, in situations where the reading capacity has been severely impaired. In T.M.'s case , it is so impaired that it takes him around 3 seconds to start to read aloud a word, when the normal time is around 200 ms. This is one of the reasons why no conclusion can be reached on the subject just by referring to neuropsychological data.

Another type of evidence for configurational features comes from studies involving special categories of words, such as acronyms (Henderson & Chard, 1976; Seymour & Jack, 1978; Besner et.al , 1979) and brand names (Gontijo, Shillcock & Kelly, 1997). These two groups of words have in common the fact that most of the time they occur in a particular type-face in the real world. However, these results are restricted to either the lexical decision or the same-different task paradigms. This issue will be discussed in more detail in the next chapter.

5.1.2 - The analytic approach

The analytics, as we have seen, argue that word recognition is done solely on the basis of the identification of the individual letters that form a string. It is also believed by the vast majority of proponents, that each letter can be recognised on the basis of an **abstract** letter identity code (ALI) that is independent of any physical features. Next, we describe the analytic approach by taking the ALI model of word recognition as the basis for that description.

5.1.2.1 - The ALI model of word recognition

The ALI theory was proposed by Evett and Humphreys (1981) who believe that the automatic access to the internal lexicon operates by activation of abstract graphemic information. At the level of the abstract letter identification units (ALIs), visual information corresponding to the letter is assimilated and account is taken of its physical specifications, such as type, case and font. The output produced is then coded independently of the specific visual form in which the letter is presented.

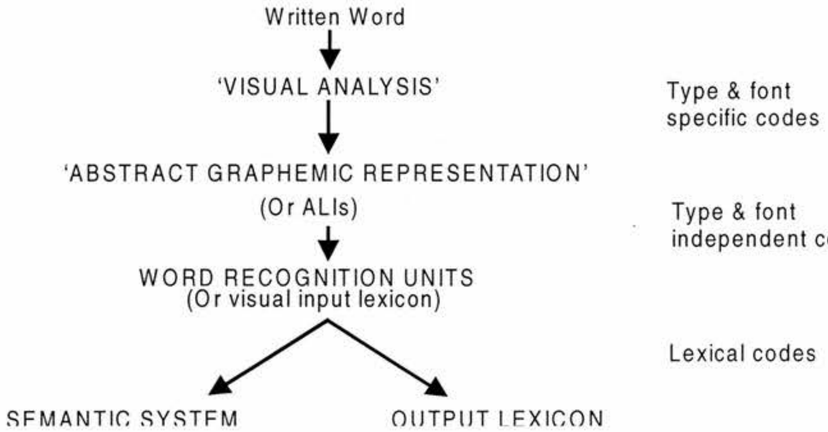

Fig.(5.2) - ALI model of visual word recognition (Howard, 1987)

It is generally held by ALI theorists, that the perception and identification of a written word takes place within a system in which there are several hierarchical levels. Identification starts, at the first level, with the extraction of a set of features from the written pattern by the visual analyser (visual analysis). It proceeds then to the next level where the access to the individual letter abstract codes is gained, prompting the graphemes to be recognised (abstract graphemic units). This way of working, where the recognition system initially extracts a relatively small number of visual features from the stimulus enables it to cope with the large number of different words that normally appear in a large variety of different typefaces. The picture below offers a more graphical idea of the ALI model.

Supporting the ALI theory are the results found by McConkie and Zola (1979). They found that eye movements to a text of words made up of letters alternating in case were not affected by changing the case of the letters during the saccades. A similar result has been reported by Rayner (1979). He presented a word in parafoveal vision and instructed subjects to make an eye movement to the stimulus, which was to be named as quickly as possible. Changing the case of the word while the saccade was being made had no effect on naming latency. Nevertheless, changes in the identity of letters in target words have marked effects on both eye movements and naming latencies (Rayner, 1975, 1979).

Another important source of evidence for the ALI model comes from the absence of the word superiority effect in tachistoscopic experiments using the alternating case paradigm. The results of these experiments have shown that any alteration to the global form of strings of letters affects the recognition of pronounceable nonwords (e.g. NORK) in the same way it affects the recognition of real words (e.g. PORK). Thus, for example, McClelland (1976) found that the word-nonword advantage was unaffected when subjects had to report as many letters as possible from briefly presented stimuli, where letters were presented in alternating case, compared to a condition where all the letters were presented in the same case. This type of result was also found in a number of other experiments, such as Besner (1983) who found that naming latency for words and nonwords was equally affected by case alternation. Adams (1979) found that the word-nonword advantage was unaffected when each letter in the stimulus was in a different size and typeface from each of the others.

Finally, work in neuropsychology with deep dyslexic patients also points in the direction of the ALI model. These patients are normally unable to perform tasks that require knowledge of either sounds or names of single letters (as opposed to T.M). However, they seem to know about the letters' abstract identities, independent of case and format as they perform successfully a range of tasks that depend only upon this type of knowledge. For example, they are capable of choosing from a lower case alphabet, which letter corresponds to H and they are also capable of judging whether pairs such as gG or gP are the same. Here I refer to the appropriate literature for details of the studies mentioned (Coltheart, Patterson & Marshall, 1980; Saffran & Marin; 1977).

To sum up, there are two schools of thought regarding the ways in which visual recognition is achieved. In the analytic point of view, the only source of information used in word recognition is the abstract identity of the individual letters forming a string. The other school of thought, the holistic, believes that not only the information about the identities of the individual letters plays a role in word recognition, but also other types of information such as the envelope and configurational features of the word.

In the next section we describe a model of visual word recognition that attempts to be faithful to the literature on visual word recognition by accommodating these two different views.

5.2 - Besner and Johnston's (1989) model

Word-shape is defined as the encoded information, in the mental lexicon, of the visual features that compose familiar words. By the same token, it is not possible for pseudowords to have word-shapes, since by definition pseudowords are strings of letters that although pronounceable have not been encountered before. This difference was seen as a means of testing the role played by shape information in word recognition. It was expected that this extra word-shape information attained by words would make them more vulnerable to alterations to their shapes than to their pseudoword counterparts. Thus, for example, changing the type-case or type-font within a word would be disruptive to the processing of configurational features, because that manipulation would render the envelope and the conjunction of contours across letters totally unfamiliar. Therefore, such disruption was expected to reduce or to abolish word superiority effects if global features had the role proposed for them. McClelland (1976) and Adams (1979) published their results with tachistoscopic experiments where they reported finding that both words and pseudowords were indeed adversely affected by case alternation. But, crucially they did not find either a reduced or abolished word superiority effect there. So, the word superiority effect in tachistoscopic report[1] is unperturbed by case alternation.

Besner and colleagues (1983) argued that a remarkable amount of weight was given to the evidence provided by McClelland & Adams in the unanimous ruling out of any role for global shape or transletter features in visual word recognition (Adams, 1979, Allport, 1979; Henderson, 1982, McClelland, 1976). They contended that the literature on reaction time had been overlooked, since it could provide examples of studies with words and nonwords with case-alternated stimuli where the word superiority effect had been affected in types of paradigms such as lexical decision and same-different matching tasks. Henderson and Chard (1976), for example, report finding a word superiority effect for acronyms such as FBI, but only when they were shown in their familiar case (i.e. upper-case). FBI could be matched faster than IBF but the same was not true for fbi vs. ibf (Henderson & Chard, 1976; Seymour & Jack, 1978).

Based on the above, they proposed the hypothesis that word recognition uses two different types of mechanism: the identification and the familiarity mechanism. The identification mechanism, the one that is normally used for reading, by definition relies on preliminary letter identification. It is utilised for tasks that require unique specification of a stimulus, as for example, tachistoscopic reports, naming latency, semantic categorisation and so on. The type of information used by the familiarity mechanism is the figural pattern of a word. It does not take into consideration information about the individual letters forming the word, it only knows them as an integral part of a visual pattern. This applies to lexical decision and same-different types of tasks.

[1] In tachistoscopic identification, subjects are shown items for very short presentation times. The experimenter records the threshold at which subjects can no longer confidently identify the items. The name is reminiscent of early days, when a piece of equipment called *a tachistoscope* was used for presenting pictures of words for very short durations.

5.2.1 - Model description

Here I will describe very briefly Besner and Johnston's (1989) model of word recognition. I will focus on a particular route, i.e., the *orthographic familiarity route* (via 6 and 7in the Fig. (5.3)), which I will use to couch our explanation for the results obtained in the experiment to be outlined next. The model shown below in Fig. (5.3), is an attempt to provide an integrated account of the hypothesis that word recognition is best seen as formed by two different types of mechanisms, the identification and the familiarity processes.

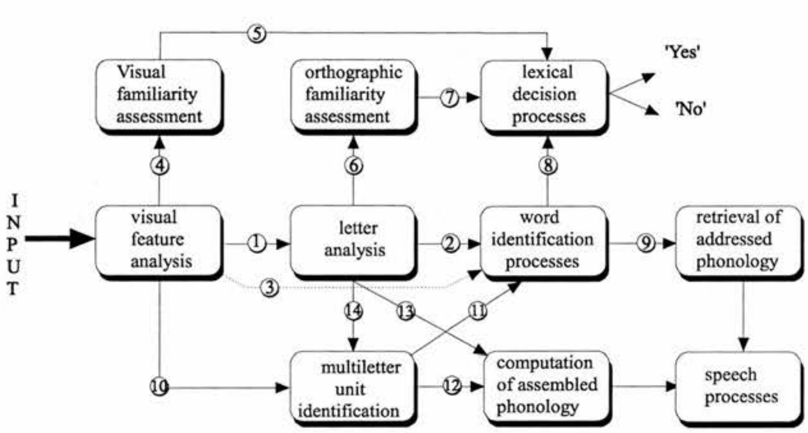

Fig. (5.3) - Besner and Johnson's model of Visual Word Recognition, 1989.

The figure above illustrates the three routes that are used to process visually presented words and nonwords. The addressed phonology route either recognises or identifies words and this route always forms words from component letters. The assembled phonology route consists of the formation of multiletter units (syllable-like units) so that words and nonwords can be pronounced. Finally, the familiarity assessment route uses item familiarity to recognise words on a lexical-decision task using whole-word stimuli as basic units of analysis. It is important to observe at this point that words however are never identified using the familiarity route. Routes 1 and 2 are those used by tasks such as naming, tachistoscopic reports, etc., where the identification mechanism must be employed. In other words, according to this model reading is done through preliminary letter identification. What is the same as to say that, this model is the equivalent of an analytic account of visual word recognition.

The hybrid nature of Besner & Johnston model is an attempt to account for the finding that there is in fact one type of situation in which case alternation produces a larger effect for words than for nonwords, and this is in responses to tasks of the lexical decision and same-different types of paradigm (Besner, 1983; Besner & McCann, 1987). This fact gave rise to the hypothesis that there might be situations for which the identification of a string is not a requirement. Assessing its pattern's degree of familiarity suffices for recognition. Besner and Johnston incorporated route 4 and 5 into their model to account for that. The model's explanation for the result mentioned above, is that because words are familiar and nonwords are not, words should produce a substantial output from the familiarity assessment mechanism, facilitating "word" responses. The assumption is that by disrupting route 4, case alternation produces a bigger effect for words than nonwords.

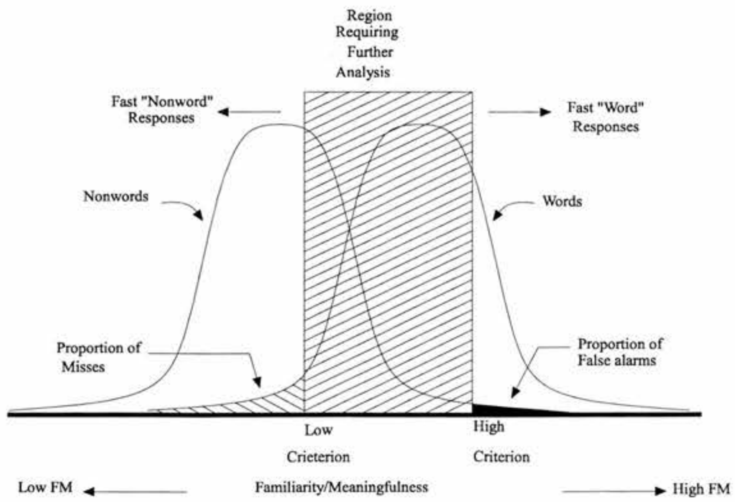

Fig. (5.4) - Balota and Chumbley's (1984) two-stage model of the lexical decision task

Route 6 (orthographic familiarity mechanism) was designed to acknowledge the possibility that a familiarity mechanism might also be activated by letter-level codes. This is a reference to Balota and Chumbley's (1984) model of the lexical decision task, depicted in Fig. (5.4) below.

Balota and Chumbley (1984) proposed that words and nonwords vary along a familiarity/meaningfulness (FM) dimension. The word and nonword distributions along this dimension are separated but overlap (e.g. the nonword *chumingly* may be more familiar and meaningful than the low-frequency word *ortolidian*). Frequency effects in the lexical decision task may be exaggerated because low-frequency words are a bit more similar to the nonwords on the FM dimension than are high frequency words. Hence, when there is insufficient information to make a fast "word" response, the subject is required to engage in an extra checking process, possibly checking the spelling of the letter string against the spelling of a word contained in the subject's lexicon.

We will not discuss here the other routes in Besner and Johnston's model, since they are not relevant to us at the moment. I refer to the original paper, i.e., Besner and Johnston (1989), for more information.

So far we have seen that there is no consensus as to which is the functional unit of word recognition. Findings in the literature, force us to assume that there are two mechanisms for word recognition and that they depend on the type of task to be performed, namely, the identification and the familiarity mechanisms. The identification mechanism is used for tasks that demand unique identification of the stimulus to be recognised, such as the naming and categorisation tasks. It uses only the information about the abstract identity of the individual letters forming the string. The other mechanism, the familiarity assessment, uses information other than the abstract identity of individual letters. We have also seen the model proposed by Besner and Johnston (1989) that aims to accommodate these two types of visual word recognition by proposing different routes of processing. We also saw that the orthographic *familiarity assessment route* (route 6) has its origin in Balota and Chumbley's seminal paper where they discuss the subject of task-specific effects' contribution to word frequency effects.

5.3 - Motivation

The influence of environmental statistics in the functional architecture of visual character recognition, even in adulthood, has recently being demonstrated by Polk and Farah (1994). A related issue is that of the contribution that upper-case and lower-case letters have in reading. There are only a restricted number of familiar ways in which case is used in the representation of a string of letters. This observation prompted us into asking a number of questions, as for example, has the cognitive visual system developed any special strategy for processing these familiar case arrangements? Also, given that, how would its processing capacities cope with naturalistic case changes?

As we have seen, the majority of the experiments on case influence, have resorted to the alternating case paradigm as a method for investigating the subject. Although, this is a very successful paradigm (Bruder, 1978; Pollatsek et.al., 1975; McClelland,1976; Adams, 1979; Mayall & Humphreys, 1996), the extreme disruption it causes to the string's appearance reduces it to a type of stimulus that is not normally encountered in the environment. Consequently, the strategies used to deal with these strings may not be the same as those used in normal processing.

In Experiment 3 a less disruptive manipulation of the stimulus appearance was chosen. An experiment was designed where the visual patterns forming the strings to be compared were those we are normally used to. The literature shows that case influence is dependent on the type of task performed. Positive results have already been found in same-different experiments designed to investigate the effects of case alternation. The goal of the experiment was to test for any behavioural difference between pairs of nonwords which exhibited different case manipulations. To test for differences between strings that are totally represented in upper-case versus those totally represented in lower-case, pairs of strings such as (beert beert) and (BEERT BEERT) were used. Also, in English, as in many other languages as well, a familiar way of signalling the beginning of a sentence is to start it with a initially capitalised word. To test for the familiarity effects involved in such situation pairs of the kind (Beert beert) and (beert Beert) were also incorporated into the experiment. Other conditions involved pairs of strings whose shapes are not familiar to readers as those just referred above, as an example, we quote the pairs (beerT beert) and (BEERT beerT). There are very few experiments that test case influence without completely disrupting the overall outline of the string (Fisher, 1975; Pollatsek et. al., 1975, Baron, 1976). However, their results are encouraging. I hypothesise that strings that exhibit a visual pattern that is closer to the ones normally found in the environment would be dealt with more efficiently. Furthermore, those having a higher frequency rate of appearance will elicit a faster response time.

The design above, has also allowed for questions concerning the arrangement of the strings forming each of the pairs. For example, the order of appearance of strings forming a pair [(Beert beert) (beert Beert)] was also addressed.

Bruder (1978), also using a same-different task found that high frequency words were considerably more affected by case alternation than low-frequency ones. He also found an advantage of pure over mixed case stimuli for nonwords that were constructed by replacing a single letter in high frequency words, and reported that the same was not evident for nonwords that had low-frequency words as their base. He controlled for the frequency of his nonwords in the traditional way, i.e., constructing them by replacing a single letter in a real, high frequency word. Here, the algorithm used in Chapter 2 is used to build the nonwords that are all composed of high-frequency onsets and codas and I resume the investigation started with in Experiment 2 of the role played by the "weirdness" variable. Our expectation is to find faster reaction time for pairs composed by nonweird strings as opposed to the weird ones.

5.4 - Experiment 3

5.4.1 - Participants

The participants were 28 volunteers from the Physics Department of Heriot-Watt University in Edinburgh. All participants had normal or corrected-to-normal vision, and each was run individually in one session of about 30 minutes. One participant was discarded due to equipment fault.

5.4.2 - Stimuli and design

A set of 280 pairs of nonwords was built to be used as stimuli in the experiment. Refer to Appendix III to see the list of them. Half of them were used as fillers. The nonwords length was either 5 or 6 letters long. They were all built according to an algorithm developed by Gontijo and Shillcock (1995). They were all *high-frequency* nonwords, in the sense that they were all formed by using a combination of high-frequency onsets and codas. Two different types of nonwords were formed: firstly, those we named "weird nonwords", i.e., those which had no nucleus + coda (or "word-body") present in the lexicon. The second type of nonwords were termed "nonweird" because their nucleus + coda combination are present in the lexicon. The first and last letters of the nonwords were controlled such that if a participant saw one word starting with a specific letter he/she also saw another nonword that finished with that same letter. This was done to avoid the artefact of specific letters interfering with the phenomenon being studied, as first or last letters will appear in upper-case for some pairs of nonwords.

The nonword pairs could be formed either by an orthographically identical set or by an orthographically different set of nonwords (fillers). There were 12 conditions for either type of pair, that is, a total of 24 different conditions were manipulated during the experiment (half of them belonging to the fillers). The variables being manipulated were the capitalisation of (one or all) the letters of the nonwords and also weirdness, i.e., the presence or not in the lexicon of the combination nucleus + coda composing them. The weirdness variable has therefore two conditions: weird and nonweird. These conditions were controlled when the 280 pairs of nonwords were created. Half of them are formed by weird and the other half by nonweird nonwords. All subjects saw a pair of nonwords only once in one of the twelve conditions.

Table (5.1) shows all the 12 conditions for the variable capitalisation, using the nonword *beert*[2] as an example.

Code names	Conditions	Code names	Conditions
A	beert-beert	B	BEERT-BEERT
C1	BEERT-beert	C2	beert-BEERT
D1	Beert-beert	D2	beert-Beert
E1	BEERT-Beert	E2	Beert-BEERT
K1	beerT-beert	K2	beert-beerT
L1	BEERT-beerT	L2	beerT-BEERT

Table (5.1) - First variable: letter capitalisation (the string "beert" is used as an example).

Each participant performed a classification task for pairs of nonwords in all 24 different conditions. The order of presentation was balanced with the use of 28 x 28 Latin Squares. The 28 Latin Squares design was used because conditions A and B, in which pairs were formed by identical strings, were treated as if each of them were 2 conditions, i.e., A1/A2 and B1/B2. This was done to keep the symmetry of the experimental design. The material was randomised for each participant[3]. Visual similarity was carefully controlled, to ensure that two nonwords forming a pair that is orthographically different did not look strikingly different from each other, since this could disrupt the effect. The same approach was used with respect to phonological similarity.

[2] Note that although the nonword "beert" can become a word (beer) when detached from its last letter, this was not the case with the other 139 words used in the experiment.

[3] The randomisation of an extensive data set and subjects was possible thanks to a program written by Stuart Boutell.

The orthographically different pairs were built to sound as phonologically close as possible.

5.4.3 - Procedure

A Macintosh computer running the PsyScope software (version 1.0.2) controlled stimulus presentation and timing. Participants were instructed that the experiment dealt with "nonwords: i.e. words that are not normally found in the language". They were asked to sit in front of the screen and rest their hands on the response box, located in front of them. They were told that firstly there would be a training set to allow them to familiarise themselves with the task. Next, the experimental trials would start. They were told that the experiment was divided into two parts, each lasting for about 10 minutes, with a break in the middle to allow them to rest.

On each trial, a fixation point appeared on the screen for 500 ms. Next, came a pair of nonwords which was displayed on the screen in a single line, to help the participant seeing each string as a unit. The stimulus was displayed using New York font, size 24 points, and bold characters. Participants were instructed to classify as fast and as accurately as possible if the pair of nonwords appearing on the screen was spelled differently or not. They were also told to disregard any other difference they might encounter. Their task was to press the button named "D" for different and "S" for same accordingly. They were also instructed to use their dominant hand to press the "D" button.

5.4.4 - Results

There are two different types of effects to be discussed. These are related to the two variables used in the experiment, namely case manipulation and weirdness. Firstly, case manipulation will be discussed.

5.4.4.1 - Case manipulation

We first present the overall results of manipulating the letter-case variable. Following that, we present the results one by one of the significant conditions. Next, we discuss these results.

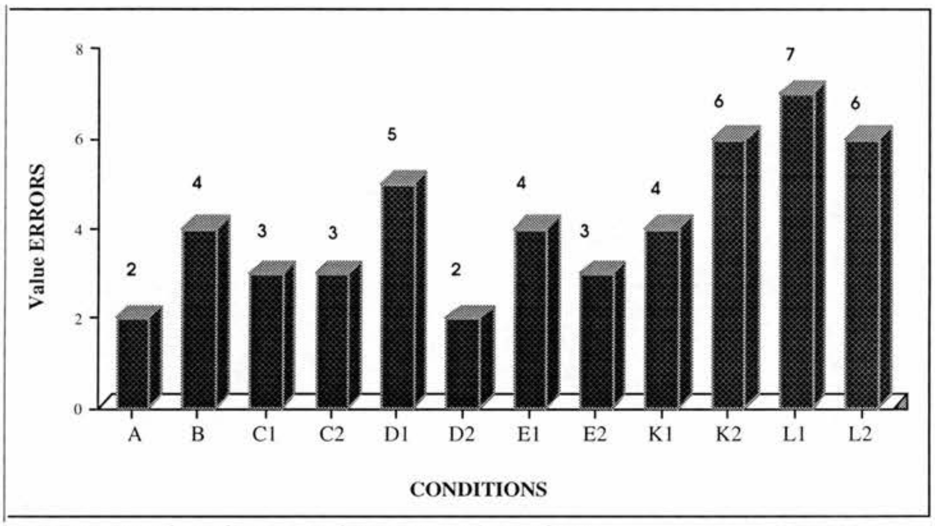

Fig.(5.5) - Number of committed errors in each condition

The total number of errors per condition, for all subjects is shown in Fig. (5.5) above.

Participants made mistakes in only a small number of trials, for example, by pressing the "S" (same) button, when the pair of nonwords was actually different. These trials were discarded from the response time (RT) analysis. The subjects average response times and their standard deviations are shown in Table (5.2) below.

Conditions	RTs	St. devs.	Conditions	RTs	St. devs.
A	1297.4	396.5	B	1361.7	429.9
C1	1497.0	460.0	C2	1482.9	440.5
D1	1391.0	407.9	D2	1449.1	483.5
E1	1503.7	436.2	E2	1452.0	400.6
K1	1523.6	447.2	K2	1483.9	419.5
L1	1541.0	466.0	L2	1574.1	423.8

Table (5.2) - Reaction Times (ms) and standard deviations for all conditions.

A one-way repeated-measures analysis of variance (ANOVA) was used to verify any statistically significant difference between the twelve different conditions. The obtained result was $F(11,26) = 12.28$ ($p < .001$). This shows that there is a statistically significant difference among conditions. A further test, i.e., a Pairwise Multiple Comparison Procedure (Student-Newman-Keuls Method) was performed to isolate the groups that differ from each other. The reaction times for each of the conditions are shown in Fig. (5.6).

As it can be observed from Fig. (5.6) below, the fastest reaction time is given by condition A (beert beert). Note also that condition A is statistically significant when compared to any of the other conditions individually. Immediately following condition A comes condition B (BEERT BEERT). Condition B is also statistically significant compared to any of the other conditions individually. Condition D1 (Beert beert) comes after condition B. Except for conditions A and B, condition D1 also presents a statistically significant difference when compared to all other conditions individually. Here, we anticipate the next results by calling attention upon the fact that by excluding conditions A, B and D1, then all the other conditions can be clustered together under the heading of *unfamiliar conditions*[1]. For, at least one of the strings composing these pairs is represented in an unfamiliar format. On the one hand they do not show any statistically significant difference in reaction time between themselves (with few exceptions that are discussed later) and on the other hand all of them present significantly longer RTs compared to conditions A, B and D1.

[1] With the exception, perhaps of condition D2, that is a complementary pair to D1.

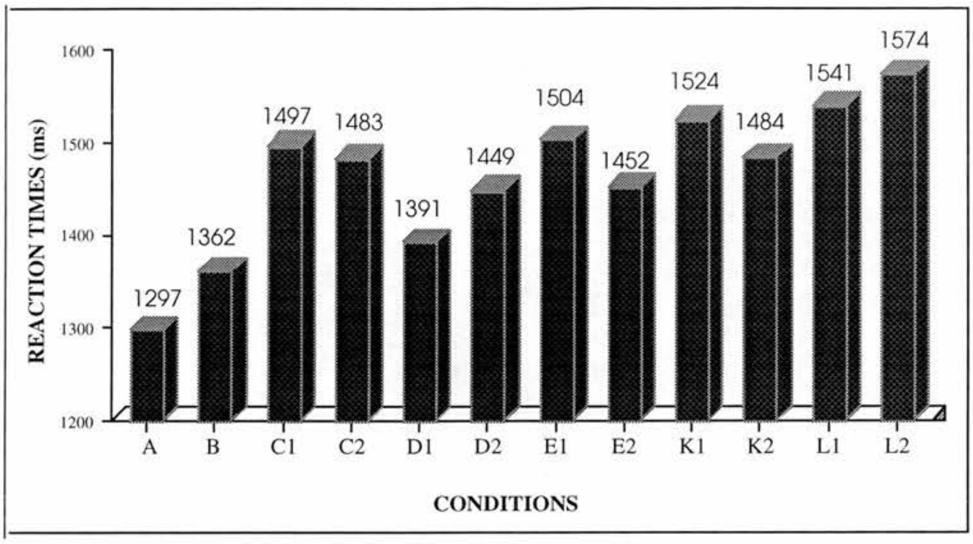

Fig (5.6) - Reaction times (ms) for each of the experimental condition

Finally, for each pair in the experiment there is a correspondent pair which has its strings inversely aligned to it in terms of a left-right orientation. For example, K1(**beerT** beert) x K2 (beert **beerT**). There is no statistically significant difference between these pairs.

The differences in reaction times (ms) between the overall conditions, as well as an indication of their level of significance can be found in Table (5.3) on the next page.

Next the results for the weirdness variable are presented. The discussion of the results above will be delayed until section (5.6).

5.4.4.2 - Weirdness

Another aspect involving familiarity that has also being investigated here is that of orthographic familiarity. As previously mentioned, a second variable manipulated in the experiment was the type of nonwords used. In the stimuli and design section, I described two different types of nonwords that were created for the experiment. For one set the nucleus + coda (word-body) is supported by the lexicon (nonweird nonwords) and for the second set the

Table 5.3 - Reaction Time (ms) differences between the overall conditions

RTs with $p < .05$ (by subjects) are marked with the symbol (♦). RTs which did not reach significant levels bear no mark.

	A	B	C1	C2	D1	D2	E1	E2	K1	K2	L1	L2
A	-	64.3♦	135.3♦	121.3♦	29.3	87.4♦	142.0♦	90.3♦	161.9♦	122.2♦	179.2♦	212.7♦
B	64.3♦	-	199.6♦	185.5♦	93.6♦	151.7♦	206.3♦	154.6♦	226.1♦	186.5♦	243.5♦	276.7♦
C1	135.3♦	199.6♦	-	14	106.0♦	47.9	6.7	45.0	26.5	13.1	43.9	77.1
C2	121.3♦	185.5♦	14	-	91.9♦	33.8	20.7	30.9	40.6	0.9	58.0	91.1
D1	29.3	93.6♦	106.0♦	91.9♦	-	58.1	112.7♦	61.0♦	132.5♦	92.8♦	149.9	183.1♦
D2	87.4♦	151.7♦	47.9	33.8	58.1	-	54.6	2.9	74.6	34.9	91.8	125.0♦
E1	142.0♦	206.3♦	6.7	20.7	112.7♦	54.6	-	51.7	19.9	19.8	37.2	70.4
E2	90.3♦	154.6♦	45.0	30.9	61.0♦	2.9	51.7	-	71.6	31.9	88.9	122.0♦
K1	161.9♦	226.1♦	26.5	40.6	132.5♦	74.6	19.9	71.6	-	39.7	17.4	50.5
K2	122.2♦	186.5♦	13.1	0.9	92.8♦	34.9	19.8	31.9	39.7	-	57.1	90.2♦
L1	179.2♦	243.2♦	43.9	58.0	149.9♦	91.8	37.2	88.9	17.4	57.1	-	33.1
L2	212.7♦	276.7♦	77.1	91.0	183.1♦	125.0	70.4	122.0	50.5	90.2	33.1	-

A= beert beert
E1=BEERT Beert

B= BEERT BEERT
E2= Beert BEERT

C1= BEERT beert
K1= beerT beert

C2= beert BEERT
K2= beert beerT

D1= Beert beert
L1= BEERT beerT

D2= beert Beert
L2= beerT BEERT

nucleus + coda combination could not be found in the lexicon (weird nonwords). The means for response time and standard deviations for weird and non-weird nonwords are presented in Table (5.4) below.

Nonword type	Means	SD
weird	1456.5	409.9
nonweird	1410.0	424.2

Table (5.4) -RTs (ms) means for weird and nonweird nonwords.

The importance of the "word-body" is supported by previous reports of its role in determining vowel pronunciation (Treisman & Zukowisk, 1988; Treisman et. al., 1995), and in mediating word-word priming (Taraban & McClelland, 1987) and consistency effects in naming (Jared et. al., 1990). See however, (Kay, 1987) and (Taraban & McClelland, 1987) for evidence of the role of onsets in nonword pronunciation. A unpaired t-test analysis shows that the matching task of nonweird nonwords was completed more quickly than those of weird words ($t = 3.546$, $df= 26$, $p <.002$). This finding is in line with the majority of research in the area and it further corroborates our hypothesis that environmental statistics play an important role in visual word recognition: Nonweird nonwords have their body in the English lexicon and, being more familiar can be matched faster.

5.4.5 - The Transformation Model

As we have seen, the use of word-shape as a source of information in visual word recognition is a controversial issue where holists and analytics take different views. Word-shape is a property of words, since it refers to the recorded information, found in the mental lexical, of a previously encountered string. Therefore, it is by definition not possible to talk about word-shape for a nonword. When the recognition of nonwords is slowed down by alternating case disruption, the consensus in the literature is that this is due to alterations of the *psychophysical distinctiveness* of the individual letters forming the string (Paap et. al., 1984). However, I suggest that this explanation is confined only to those cases where the familiar aspect of the string has suffered drastic changes in its configuration. It does not apply, for example, to account for the results obtained in the current experiment. Care was taken in the experiment, to have the physical appearance of the strings manipulated in such a way that the majority of them replicate the printed patterns commonly found in thc environment. I propose that a more elegant explanation for the phenomenon described above can be given at the level of the mental representation of the abstract letter identities by using the *orthographic familiarity route* of Besner and Johnston's (1989) model. This route admits the familiarity contamination for tasks where familiarity effects are implicated. See the Fig. (5.7).

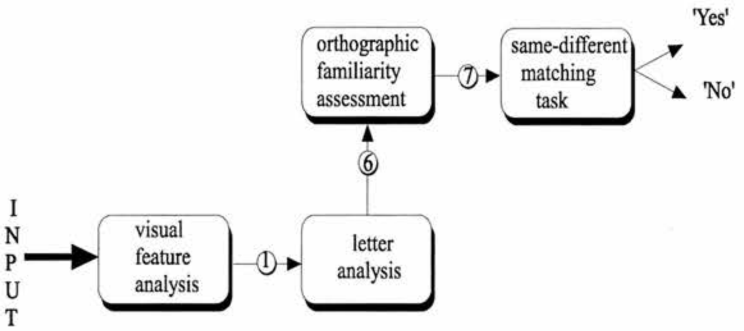

Fig. (5.7) - The orthographic familiarity assessment route adapted for the same-different task.

As seen previously, the orthographic familiarity route was designed to acknowledge Balota & Chumbley's (1984) findings, in which the possibility that a familiarity mechanism might also be activated by letter-level codes is entertained. Here I propose that not only orthographic familiarity is accessed by this route but also other types of visual familiarity, such as familiarity of case. For that reason, I propose here to rename the *orthographic familiarity route* as the *visual familiarity route*. Next, I attempt an explanation of how its mechanisms work.

I suggest that to reach the abstract code of a letter, its physical representation in the world must undergo a transformational step. The ease with which this transformation can be accomplished is linked to the frequency with which that type of physical letter input is found in the world. A metaphor illustrated by Fig. (5.8) might be of use here.

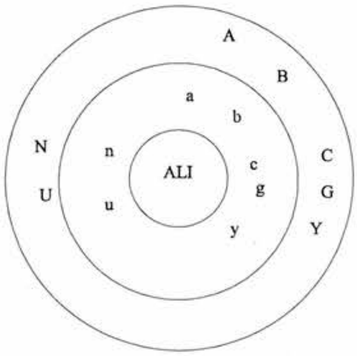

Fig.(5.8) - Scheme of the relationship between the ALI's representation and their physical counterparts.

Consider the concentric circles above. They represent an abstract space and the innermost circle is where the abstract letter codes reside. Consider a hierarchical arrangement of successive layers surrounding the central circle. In each of the layers there is a different physical representation of individual letters. The hierarchy is such that those physical attributes that are more frequently encountered by the reader occupy the layers nearest to the centre. Also, let us assume that the nearer a representation is to the centre, the easier (i.e., the faster) it becomes to transform it into its abstract counterpart. Evidence has been established linking the ease of processing and the frequency with which a stimulus is encountered in the environment by Polk and Farah (1994, 1996). An account of this evidence has been given in section 2.2. of Chapter 2. To recap the above: on the one hand, the nearer a physical representation is to the centre the faster it is processed. On the other hand, the further away, it finds itself, the slower its processing becomes.

Let us now suppose that the lower-case letter representation layer is the one occupying the space immediately outside the central circle. In the next layer lies the representation for the upper-case letters. There should be other layers, but it suffices to concentrate on these two. The transformational model predicts that for tasks where the contamination of familiarity obtains, as for example, in lexical decision and same-different tasks, subjects' performance should be susceptible to case influence. A slightly better performance should be obtained for stimuli presented in complete lower-case as compared to those presented in complete upper-case letters. In fact, supporting evidence can be found in the literature, even for naming experiments, such as reported by Baron and Strawson (1976), who found that reading times were marginally slower for upper- than lower-case words. Also, Fisher (1975), shows a reading time advantage for people reading, a *normal* paragraph[2], compared to a paragraph that was presented written only in upper-case letters.

Another feature of this model is that of the computational load involved in case switching. To use the metaphor of layers again, this implies, a penalty in terms of time when the processing of a string demands switching between the different layers of the circle. In other words, it takes longer to recognise a string that is not represented in consistent-case, i.e., a mixed-case string, which is represented by a mixing of both upper- and lower-case letters.

However, the theory also acknowledges the role played by the position that an upper-case letter appears in a string. In all models of visual word recognition, the ORDER of the letters forming a string matters for its encoding. I suggest that the POSITION where a capitalised letter appears in the string is equally important. Further, I propose that in strings where case is not consistent, only INITIAL capitalisation is seen as familiar and thus it is the most easily (fastest) processed. The main reason for this might reside in the historical background of the structure of English and the European languages in general (as all of them use the roman alphabet). An initial capital letter is used in two very frequent situations in these languages, namely, starting a sentence and in proper names to refer to something or someone. It is expected that any word at the beginning of a sentence will have an initial capital letter, regardless of the word class.

[2] Fisher called a "normal paragraph", the one predominantly written in lower-case, except for the word at its start and the proper names it contained, that were all initially capitalised.

A small statistical study of printed text was conducted by us as a means of having some quantification of the above. The distribution of letter cases for one chapter of each of two different novels were counted. The results are shown below in Table (5.5).

Novel's titles	Total no. of words	No. of lower-case words	No. of initial upper-case words	Proper Names
The Moon and Sixpence by Somerset Maugham	5920	5365	447	108
The Devil's alternative by Frederick Forsyth	6120	5512	418	182
Totals	12040	10317	865	294
Total Percentages	100%	90.3%	7.3%	2.4%

Table (5.5)- Distribution of letter-case in novels written in English.

As can be seen, the percentage of capital letters appearing in the text that are not proper names is more than double. These letters always appear at the beginning of sentences. We hypothesise that this situation leads the cognitive system to become trained in ignoring the initial capitalisation of ordinary common nouns encountered at the beginning of sentences. The effects of this state of affairs can be seen in people's prompt reaction when recognising the similarity of spelling between two strings of letters that differ only with respect to initial capitalisation.

Now that the foundations for the transformation model hypothesis have been laid, let us turn to the set of predictions it makes. First, it predicts that there is a hierarchical structure attached to the easiness of string processing. The strings composed of consistent-case letters are easier to process than those formed by mixed-case letters.

Let us consider first the consistent-case strings. These can be composed of either lower- or upper-case letters. According to the hypothesis being put forward here, the strings composed by the lower-case letters should be the easiest type to process. The reason is that they would be located immediately outside the core of the circle. This privileged position is a function of the frequency with which these types of letters are found in the environment. The next easier to process are the consistent upper-case strings that are located in the adjacent layer. Summing up in more general terms: the less frequent a physical representation of a letter is encountered the further is its location from the innermost part of the circle and the longer it takes to process it.

The hierarchy involved with the mixed-case strings is more complex than that applied to the consistent ones. Recall that the switching between different cases is penalised during processing. An extreme example of mixed-case strings is the stimuli used in aLtErNaTeD cAsE experiments. The theory here predicts that these are the ones that would take the longest time to be processed. In other words, the more case switching there is in a string the longer it would take for the cognitive system to process it.

In a somewhat similar fashion, Besner (1983) has raised the issue that the unfulfilled expectation regarding consistency plays a role in slowing down the process of recognition. He pointed out that one of the explanations for the fact that case alternation impairs performance in naming latency task is that it simply reflect s an increase in uncertainty: one out of 52 alternatives for a case alternated letter as opposed to one out of 26 for a letter whose case is already known.

A different situation is that of strings that are only initially capitalised. Recall that initial capitalisation holds a special familiarity status due to its frequent appearance in words at the beginning of sentences and in words belonging to category of proper names. Precisely because of the familiarity issue our model predicts their processing to be as quick or at least as equal as to that of the consistent upper-case ones. However, the model predicts that longer time should be taken for processing strings where the inconsistent-case letter is not located at the initial familiar position but elsewhere in the string, for instance, at its end.

5.4.5.1 Mechanism: Quantitative analysis

Next, I offer a quantitative account of the transformation mechanism in terms of how many "units of time" are required to recognise strings that bear different case patterns. I start by first explaining how the mechanism works for single word analysis and second how it accounts for the task of matching pairs of strings as is required from subjects in the same-different matching task paradigm.

A) Single string recognition

1) - A basic assertion of the transformation model is that lower-case letters take a shorter time to be recognised than capital letters. Here we assume that lower-case letters will take one unit of time to be recognised ($u=1$) whereas capital letters will take one and a half time units ($U = 1.5$). Thus based solely on letter case and ignoring all other features, the string "beert" would take 5 time units to be recognised and the string "BEERT" would take 7.5 units.

2) - Letter case consistency (C) within a string was also pointed out to be important. If the string has mixed case letters there will be a penalty in time of recognition to be added to the time spent in the instances reported in paragraph (1) above. This penalty is accounted for by taking twice the time of the inconsistent letter. The consistency C of a single string can therefore be calculated by adding the times obtained in (1) and (2). For example, C(beert) = 5; C(BEERT) = 7.5; C(Beert) = (4 + (2 x 1.5)) = 7 and C(bEERT) = ((4 x 1.5) + (2 x 1)) = 8.

3) - Also of significance to the process is the position of the capital letter (**PCL**). If there are upper and lower case letters in a string, the position of occurrence of the capital letter(s) will affect the recognition time in the following way. If the capital letter appears at the beginning of the word, it is simply equal to one and a half time units (**PCL** = 1.5). On the other hand, for every other positions of the word, the penalty incurred will be ten times the value above. Thus **PCl**(Beert) = 1.5; **PCL**(beerT) = 15; **PCL** (bEeRt) = 30;

PCL(BeErT) = 31.5. Note also that PCL(BEERT) = 0, because there are no mixed case letters in this instance.

B) Matching pairs of strings

When two strings are being compared, if the corresponding characters are not in the same case, another penalty will be incurred. This gives rise to the final mechanism below:

1) - Consistency between strings (CBS)

This is calculated by comparing a pair of strings. For every pair of inconsistent characters, 2.5 units of time (u = 1 + U= 1.5) is added to CBS. Thus, CBS(beert Beert) = 2.5 ; CBS(BEERT Beert) = 4 x 2.5 = 10

Using the mechanism described in A and B above, it is now a simple matter to calculate the predicted recognition time for different pairs of nonwords. The total recognition time (T) will be given by the sum of the consistency time (C), the position of capital letters (PCL) and the consistency between words (CBS):

Total = C + PCL + CBS

Various examples of this calculation are given below.

beert beert C = 10
PCL = 0
CBS = 0
=> Total = 10

BEERT BEERT C = 15
PCL = 0
CBS = 0
=> Total = 15

Beert beert $C = 7 + 2 \times 1.5 = 10$

$PCL = 1.5$

$CBS = 1 + 1.5$

=> Total = 13

Beert BEERT $C = 4 + 3 + 5 \times 1.5 = 14.5$

$PCL = 1.5$

$CBS = (1 + 1.5) \times 4 = 10$

=> Total = 26

beert BEERT $C = 5 + 7.5 = 12.5$

$PCL = 0$

$CBS = 5 \times (1 + 1.5) = 12.5$

=> Total = 25

beert beerT $C = 5 + 4 + 2 \times 1.5 = 12$

$PCL = 15$

$CBS = 2.5$

=> Total = 29.5

beerT BEERT$C = 4 + 2 \times 1.5 + 5 \times 1.5 = 14.5$

$PCL = 15$

$CBS = (1.5 + 1) \times 4 = 10$

=> Total = 39.5

BeErT - bEeRt $C = (2 \times 1.5 + 1 + 2 \times 1.5 + 1 + 2 \times 1.5) + (1 + 2 \times 1.5 + 1 + 2 \times 1.5 + 1) = 20$

$PCL = (1.5 + 15 + 15) + (15 + 15) = 61.5$

$CBS = (1.5 + 1) \times 5 = 12.5$

=> Total = 94

It is clear that the examples 1-8 are in increasing order of recognition time. These predictions can be compared with the experimental results obtained, if a small simplification is made. The statistical analysis of the results showed that there is no statistically significant RT differences for conditions (Beert-beert/beert-Beert, BEERT-Beert/Beert-BEERT, BEERT-beert/beert-BEERT, beerT-beert/beert-beerT, BEERT-beerT/beerT-BEERT). It is therefore reasonable to cluster these pairs of conditions together reducing from 12 to 7 different cases. These are shown below in Table (5.5) in increasing order of Rts

Conditions	Example	RTs (ms)
A	beert-beert	1279.9
B	BEERT-BEERT	1361
D1/D2	Beert-beert/beert-Beert	1391/1449.1
E2/E1	Beert-BEERT/BEERT-Beert	1452/1503.7
C2/C1	beert-BEERT/BEERT-beert	1482.9/1497
K2/K1	beert-beerT/beerT-beert	1483.9/1523.6
L1/L2	BEERT-beerT/beerT-BEERT	1541/1574

Table (5.5) - Pairs of conditions in increasing order of Reaction time.

Comparing the experimental results above, with the examples 1-7 (calculated according to our model), they are in perfect agreement. It is also interesting to compare examples 4 and 7. There is a large predicted difference in recognition time $\Delta RT = 122.1$ ms.

Note finally the highly disruptive alternating case presented as example in 8 above. There is a very large penalty due to the position of capital letters (PCL) and also another large penalty for consistency between words (CBS).

This results in $T = 94$, i.e., nine times longer than the recognition time for the simplest case (beert beert). Our proposed model can therefore quantify the possible mechanisms involved in the comparison of two strings in a same/different matching task. Furthermore, the results predicted by the model are in perfect agreement with the experimental material.

Summarising, the predictions of the transformational model are as follows: consistent-case strings should be processed faster than mixed-case ones. For consistent-strings, those composed of lower-case letters should be easier to process than the upper-case ones. With regards to mixed-case strings, the slowest processing is reserved for strings that are entirely alternated. Those with only one inconsistent letter should take less time to process than those that possess more then one alternated letter (although a matter of sensitivity might prove to be an obstacle in testing this hypothesis experimentally). Finally, there is the phenomenon that inconsistent strings with the familiar initial capitalisation are processed as fast as the consistent strings.

5.4.6 - Discussion of the experimental results

The predictions of the transformational hypothesis model were tested by the current experiment. Let us start by looking first into the consistent-case conditions, that is **A** (beert beert) and **B** (BEERT BEERT). As predicted by the model we find that **A** is the fastest condition. As we have seen the lower-case representation is the one that stands closest to the abstract identity of the letters it represents.

If we move now to the mixed-case conditions, we are, according to the transformational hypothesis model, to expect **D1** (**B**eert **b**eert) and **D2** (**b**eert **B**eert), respectively, to be the next fastest conditions. In fact this is exactly the result we have obtained. As remarked previously, although belonging to the category of mixed-case strings, their inconsistent letter is located at the familiar initial position.

Finally, we are left with the rest of the pairs, that is E1/E2, K1/K2 and L1/L2. These pairs take significantly longer to process than any of the other pairs mentioned before. However, there are no significant differences between them with the exception of the pairs L2 x E2 and L2 x K2. Next we discuss these two pairs. Let us take first the pairs L2 (beerT BEERT) x K2 (beert beerT). As predicted by our theory it takes longer to match the pair L2 than the pair K2. The difference between the two pairs lies in the fact that in L2 one of the strings is represented in consistent upper-case (BEERT) and in K2 the string is represented in consistent lower-case (beert). As we saw, the time taken to process a consistent lower-case string is shorter than that to process a consistent upper-case string.

A similar interpretation can be also given for the significant differences between the pairs L2 (beerT BEERT) x E2 (Beert BEERT). As one would expect it takes longer to process L2, since the only two strings that differ between the pairs are the strings "beerT" and "Beert" and corroborating what we have seen so far, it is easier to process a mixed-case string where the only capitalised letter is at the familiar initial position.

This section can be concluded with the remark that the model described in this chapter is able to explain all the results obtained in the current experiment.

5.5 - Conclusions

The experimental results presented in this chapter have shown that even mild case changes in a string of letters can cause disruption to the process involved in its recognition.

It has been suggested that familiarity effects can be activated by letter-level codes (Balota & Chumbley, 1984). Based on this suggestion, Besner and Jonhson (1989) had incorporated into their model of word recognition the "orthographic familiarity route". Here, we take a step further and suggest the mechanisms by which this route works, i.e., our transformation model. Note however, that the concept "orthographic" is being interpreted not only in the strict sense of assessing a familiar arrangement of letters in a string but also in terms of the visual aspect of the letters composing the string, here more specifically, in terms of case. This hypothesis predicts that there is a hierarchical structure concerning the easiness with which strings are recognised by the cognitive system. The easy with which a string is recognised is dependent upon the level of disturbance that has been caused to its physical appearance and how much this change causes it to depart from its more familiar shape. Our model successfully explains the intricate pattern of results obtained in the current experiment. It can also be used to explain other results in the literature. For example, as already mentioned in section (5.3), the results obtained by alternated case experiments (Mayall & Humphreys, 1996). Therefore there is strong experimental evidence showing that the "transformation model" is a viable explanation for some of the familiarity effects encountered in visual word recognition.

In this chapter we have also investigated orthographic familiarity in its more restricted sense of an acquaintance with groups of letters that are encountered more frequently. Our results have demonstrated that nonweird nonwords are processed faster than the weird ones. The terminology weird vs. nonweird was created by us to distinguish between nonwords which were composed by a word-body (nucleus + coda) that is present in the lexicon, from those where the combination of (nucleus + coda) could not be found in the lexicon. As a general conclusion, our results are a further confirmation of the importance of the statistics of the environment on visual word recognition.

Postscript: Revisiting a neuropsychological case

A frequent effect of acquired brain damaged is the disorder of visual hemineglect (Robertson & Marshall, 1993; Copland & Moscovitch, 1987; Riddoch & Humphreys, 1987). Its essential feature is that patients tend to ignore stimulation contralateral to the site of the lesion. Instead of being a single entity, this condition can be fractionated into a number of discrete syndromes (Riddoch, 1990). One of the disorders often observed in patients with visual neglect is known as *neglect dyslexia*. In this disorder, reading is compromised by frequent errors affecting the contralesional portion of either words or text (Arguin & Bub, 1992; Caramazza & Hillis, 1990; Patterson & Wilson, 1990; Riddoch, Humphreys, Cleton & Fery, 1990).

In this postscript we revisit a striking neuropsychological case described by Patterson and Wilson (1990) of a patient (TB) with a posterior left-hemisphere lesion. He performed within the normal range on tests for visuospatial neglect, but demonstrated a surprisingly discrete deficit in initial letter processing of visually presented words, typically making mistakes such as reporting "nose" as "rose", for instance. These data are important because they potentially provide insight into the relationship between visual processing and linguistic processing; do position-specific letter "slots" exist independently from positional information in general visual processing?

Our goal here is to reinterpret TB's performance in the light of the findings of the present chapter, where it has been shown that the precise orthographic representation of initial letters is more labile than that of other positions. We start by describing TB's deficit, then we review some of the research on the status of the initial letter. Next, we discuss the relationship between the issues of capitalisation and initial position and finally we present our conclusions.

P.1 - TB's deficit

Patterson and Wilson report the results of a comprehensive series of experiments, examining TB's impairment along a number of dimensions, which we now summarize. TB's reports of written words and nonwords were typically either correct or differed from the target by only the initial letter: "sit" was reported as "kit", "light" was reported as "right", although he also made errors in which extra letters were added to the beginning of the word, as in "pout" being reported as "spout". Patterson and Wilson describe TB as "one of the purest, most dramatic versions of the word-initial deficits thus far described in the literature" (p. 452). TB was better at identifying upper-case words and isolated letters, compared with lower-case, but this was apparently not due simply to a size difference. His individual lower-case letter naming was at 58%, upper-case at 77%. No particular letters seemed to be reliably difficult. His deficit also extended to random letter-strings and to mixed letter/number strings like "3c8n5e" and "a6s5n3". Patterson and Wilson rule out an interpretation in terms of the attentional dyslexia described by Shallice and Warrington (1977) on the grounds that, although he performed better with mixed strings like "b74k" than with random letter strings like "btqk", the errors produced did not resemble intrusion errors from the rest of the string. Placing elements abutting the left end of the word, like "2land", "xland" and "bland" marginally improved performance, but it was not until the element to the left constituted an actual word, as in "cashland" or "darkfold", that he performed substantially better. His deficit was also in evidence when the task called for the words to be defined: for instance, he defined "bead" as "A strap for holding a dog". In identifying words from oral spelling, he has no impairment, but Patterson and Wilson raise the possibility that his impairment was also manifest in writing. When nonstandard stimuli, such as vertical or mirror-reversed words, and nonstandard responses, such as right-to-left naming, were investigated TB's performance deteriorated markedly - making Patterson and Wilson question the interpretability of these particular data - but showed a more generalized deficit rather than the discrete initial-letter impairment found for standard stimuli and responses.

Patterson and Wilson discuss the data with reference to the Interactive-Activation Model of visual word recognition (McClelland & Rumelhart, 1981). In these terms, they argue that TB's condition might be parsimoniously described as a discrete deficit of letter recognition units at position 1, with no other aspects of the model being impaired. They present five observations that support this interpretation, and discuss seven potential challenges to the claim. We return to these individual arguments later in the discussion. Patterson and Wilson conclude their exploration of TB's case with the claim that comparable evidence for specific word-medial letter-position slots is unlikely to be found, but that a final letter version of TB's impairment might not be especially surprising.

Can any account be given of isolated word reading that predicts substantially greater- or even unique - vulnerability of the initial letter over the other letters? If no such account is forthcoming, then the behavior of TB stands as evidence for the representation of letter position in "slot" terms, with the implication that if the initial letter can be thought of as occupying a slot, then perhaps the rest of the letter positions may be characterized as slots too. We next review some of the studies of isolated processing that have suggested a different role for the initial letter compared with the medial letters. The critical data necessary to account for TB's behavior must involve the initial letter being represented less securely, or being informationally less salient, than other letters. We will see that the studies suggest just the opposite state of affairs.

P.2 - The Status of the initial letter

The initial letter position in English words is the most important letter in terms of information. Yannakoudakis and Hutton (1992) present quantitative evidence of this effect in speech, showing that redundancy increases overall across the different segment positions, with the initial position being more informative; this feature of the structure of spoken English words is reflected in the orthography. For monosyllabic words, this informativeness partly reflects the different availability of consonants and vowels in (C)(C)(C) V(C)(C)(C) structures. Rumelhart and McClelland (1981) describe the processing effects of this U-shaped information curve across letter positions in monosyllabic words in the Interactive-Activation Models as resembling processing from the outside in (p. 76).

Jordan (1990) reports experiments in which the outside letters of monosyllabic English words are briefly visually presented on a screen. When accompanied by a minimal amount of visual stimulation between the two letters, such stimuli are preferentially processed, compared with the interior letters of words.

In studies that make use of degraded stimulus conditions, the end letters are typically reported as being recognized before the medial ones. Shillcock and Kelly (in press) described just such results arising from the Clarity Gating technique, in which blurred, isolated words presented on a screen become progressively clearer, with subjects reporting the identity of the word at each stage. In their experiments, the last letter was typically identified first, in the more blurred conditions, followed by the first letter, with the medial letters being recognized last. The fact that exterior letters are bounded by white space on one side makes them more readily perceivable, but is unlikely to be the only reason for the special processing they receive.

In summary, from these few examples of studies involving isolated words it is clear that the initial letter-position of monosyllabic English words is typically uniquely salient, processed more quickly, and generally critical for the identification of the word. There is no evidence of a qualitatively less secure representation of the initial letter, compared with the medial letters, and no suggestion that the initial letter-position might be more vulnerable to impairment than the other letter positions. The data in this section support Patterson and Wilson's account of the data as idiosyncratic damage that happens to have selectively impaired a single slot in a general lexical template. If there is no evidence for the vulnerability of the initial position in terms of lexical processing, then the initial position must have been impaired because it is at a spatial extremity. Patterson and Wilson explore the interpretation that TB's behavior might be characterized as attentional dyslexia (Shallice & Warrington, 1977) or positional dyslexia (Katz & Sevush, 1989), but are discouraged from accepting this interpretation because of the absence in TB's responses of intrusion errors from the rest of the stimulus.

TB's deficit is ipsilesional. Although, as pointed out above, left neglect dyslexia might more typically be expected to occur contralesionally, Costello and Warrington (1987) show that a dissociation between visuospatial neglect and neglect dyslexia is possible. Nevertheless, the location of TB's lesion militates against a more general attentional explanation and in favor of a specifically linguistic one, if any such explanation were forthcoming. We turn now to a consideration of the real-world pressures on the processing of the initial letter position in English words and suggest a processing explanation of TB's data.

P.3 - Initial letters and capitalisation

As Patterson and Wilson remark, the initial letter of a string is both visually and orthographically distinctive (p. 475). One aspect of this distinctiveness, as mentioned previously, is that English words frequently begin with a capital letter, either because the word is a name or for punctuation reasons, in that the word begins a sentence. Table 4.12, in the previous chapter showed the distribution of capital letters for one chapter of each of two different novels. Although commercial/public text - brand names, directions, etc. - show greater variety in all aspects of typography (and include a larger proportion of completely upper-case text) most text encountered by adult readers conforms to the distribution shown in table 4.12.

Initial capitalisation at the beginning of a sentence has a different function than that of cueing the reader to the status of proper name. Almost any category of word can be seen initially capitalised at the starting of a sentence. Thus, the reader is obliged to cope with the irrelevant variety in the orthography of the initial letter-position by disregarding the initial capitalisation of the word as a signal of its categorical status. We may expect such real-world considerations to find an expression in the predispositions of the adult processor. A paradigmatic example of such sensitivity to the informational demands of real-world reading is found in a study reported by Polk and Farah (1996). Canadian postal workers who dealt extensively with postal codes consisting of mixed strings of letters and numbers (e.g. EH8 9LW) were found to have a reduced "pop-out" effect for letters embedded in short strings of numbers, compared with matched controls.

From the real-world facts of capitalisation, we might expect adult English readers to show reduced effects of orthographic information concerned with case in initial letter-position compared with all other letter-positions. They should show greater facility in converting the precise orthographic form into some more abstract, graphemic representation. This prediction is confirmed by the results obtained in our previous experiment, where the reaction times for the three most familiar conditions B (beert beert), A (BEERT BEERT) and D1(Beert beert)/ D2 (beert Beert) emerged clustered, being responded to, significantly faster than all the other conditions. In short, subjects were adept at ignoring specific orthographic information concerning the initial letter - its case could be rapidly and effectively ignored - as is desirable in real-world reading. This study provides us with an effect that applies only to initial letters and that involves the orthographic representation of these letters being less secure or less enduring.

Consider the range of uppercase fonts: A, *A*, A, A, A, A, A. There is no simple algorithm for replacing an upper-case instance with lower-case equivalent, and in any case this would be redundant once the letter had been converted into an abstract graphemic representation. This facility fits the bill as a putative account of TB's impairment, given that it uniquely affects initial letters of all words, and involves a less secure representation of the orthographic details of the initial letter.

P.4 - Discussion

We hypothesize that TB's impairment is the result of a general impairment in the capacity to produce/retain/manipulate abstract graphemic information. The fact that this general impairment has surprisingly discrete effects on the initial letters of words is the result of first, the level of the deficit - it is not so severe as to interfere with most of visual word recognition - and second, a processing predisposition based on the real-world requirement to abstract quickly away from orthography of the initial letter. TB has difficulty in sustaining a graphemic representation and relies more on earlier orthographic levels of representation; this produces trouble in integrating the initial letter with the rest of the word, in that TB has effectively traded the orthographic details of the initial letter for a graphemic representation, and only imperfect details of the orthography of the initial letter may be retrieved. We now itemize several aspects of TB's behavior that support this interpretation.

1. As is the case with the majority of neglect dyslexia patients, the majority of errors committed by TB (86%) are substitutions (fright -> bright) rather than deletion errors (trail -> rail) [1]. This seems to be an evidence that although TB is aware of the correct length of the string, he has problems in abstracting correctly the identity of the letter which should occupy the initial slot of that string. As we have pointed out, this might be due to the fact that the letter occupying the initial position in a string might require more flexibility of processing as a result of real-world demand in converting between upper-case and lower-case printing.

2. TB performs better (although poorly) with upper-case over lower-case letters. The same upper-case advantage applies to nonwords as well. It is not difficult to see that this inability in performing well and at the same level with upper- and lower-case strings can pose some hindrance to TB's ability in processing the initial letter of a string, since an efficient conversion from one case to the other is expected in the processing of the first letter of a string.

3. A further interesting aspect of TB's neglect errors is that they are not intrusion errors, as tend to be the case for attentional dyslexics. The problem

[1] Arguin & Bub (1990) report the case of EB that is an exception in this respect.

does not seem to be originating in interference coming from other letters constituting the string.
4. Finally, it should be observed that although the account offered here is not aimed at explaining all the phenomena surrounding TB's case, it does not conflict with some of the best accepted explanations that are given in the literature corresponding to those phenomena. For example, the fact that (as is the case with the majority of neglect patients) TB commits fewer neglect errors with words than nonwords. Probably, the two best accepted accounts of this constraint are those of Patterson and Wilson (1990) and Caramazza and Hillis (1990). According to the former, neglect dyslexics are merely inferring or guessing the neglected portion of a word using the orthographic constraints of written language. For Caramazza and Hillis the accurate encoding of the letter identities of the non-neglected portion of a word may be sufficient to address its internal lexical representation adequately and thereby to recognize it correctly. In short, the explanation given here of TB's problem with the initial letter strings can be seen as additional to theories already existent.

In conclusion, we suggest that Patterson and Wilson's patient was not idiosyncratically impaired in the first letter slot of a general visual word template, composed for one slot per letter. Rather, the impairment was a general impairment in sustaining visually presented words at a level of representation more abstract than orthographic. The orthographic representations of initial letters are inherently more vulnerable due to the need to cope with capitalisation in normal reading. Hence initial letters may be omitted in reports of single words.

Chapter 6

Brand names and their contribution to visual word recognition

After succeeding in show that initial capitalisation influences subject's performance with nonwords in Chapter 4, the next step was to investigate the impact of other familiar and unfamiliar patterns of capitalisation (Chapter 5) in visual word recognition. The more familiar a capitalisation pattern is the faster the latencies obtained using the same-different matching paradigm. Above, not matter what type of tasks positive effects were obtained with capitalisation. Now, I examine the hypothesis that familiarity effects in visual word recognition are affected by the nature of the task performed (Besner et. al., 1984), by extending previous work done with acronyms to a new category of words, i.e., brand names. I start by discussing previous experiments, using acronyms, that were designed to investigate the role of familiarity in visual word recognition. Next, I discuss some of the motivation behind the new set of experiments being presented here and describe the experiments in detail and also their analysis. Finally, the results are discussed in the light of Besner and colleagues hypothesis.

6.1 - Background and general overview

During the seventies, the word superiority effect (WSE), that is, the phenomenon that words are perceived faster than nonwords was intensely debated (Baron, 1975; Barron & Pittenger, 1974; Egeth & Blecker, 1971; Eichelman, 1970; Henderson, 1974). One of the reasons why it produced so much interest was that the mechanisms underlying the effect were believe to be those normals used to facilitate reading. The experiments investigating the role of familiarity in perception that are reported below are part of the effort carried out during the seventies to understand those mechanisms.

In the present context familiarity is defined in relation to the concept of word-shape. In its turn, word-shape is defined as the encoded information in the mental lexicon of the visual features that compose familiar words. The more frequently a string is encountered in the environment, the more familiar its visual features become.

6.2 - Acronyms' role in earlier experiments

6.2.1 - Henderson (1974)

There are at least three dimensions that have been considered when looking for an explanation for the WSE, namely meaning, familiarity and the orthographic structure of words. However, one of the prevalent early views was that the knowledge of the spelling regularities of language (i.e., the structure of words) was a sufficient condition for the WSE (Baron & Thurston, 1973; Barron & Pittinger, 1974). Reicher (1969) and Wheeler (1970), for example, have shown that participants do better at deciding which of two letters has been presented tachistoscopically when the critical letter appears as part of a word, than alone or as part of a scrambled (non)word. This facilitation, it was argued, was propitiated by the fact that the transition between letters in a word is redundant (transitional letter redundancy view). Baron & Thurston (1973), specifically excluded meaning as a factor, since they found spelling regularities to be a sufficient cause of the WSE.

Thus, Henderson (1974) decided to investigate whether meaningfulness could also be a locus for the WSE. He did so by running a binary classification task experiment (better known as the same-different task) where half of the stimuli were highly familiar acronyms (like FBI) and the other half invented strings (like IFB). The crucial element of the design was the independence of orthographic regularity from meaningfulness. He reasoned that, although acronyms are part of the mental lexicon, they do not follow the spelling regularities of English, i.e., acronyms are not regulated by the spelling rules of English. For example, in the case of *FBI*, the letter *F* is followed by the letter *B;* according to the spelling rules of English, the consonants *F* and *B* never immediately follow one after the other. Therefore, any superiority effect obtained for the meaningful type stimuli (FBI) over the meaningless one (IBF), can not be explained in terms of orthographic factors but instead, it can only be due to a meaning factor. In fact, a WSE effect was found and the conclusion that lexical membership is facilitatory in the absence of orthographic regularity (FBI vs. IBF) followed. Support for this finding, that lexicality influences the WSE independently from orthographic regularity, also come from two other sources. First, real words are superior to orthographically matched pronounceable nonwords in threshold experiments (Manelis, 1974), and also RT (Barron & Pittinger, 1974) experiments. Secondly, there is a positive relationship between word frequency and speed of response, in experiments involving visual comparison, that is, the more frequent a word is, the faster it is recognised (Chambers & Foster, 1975).

The remarkable interest in the WSE had its source in the belief that the clarification of its mechanisms could lead to a better understanding of the processing of reading itself. Thus, the locus of the WSE was a hotly debated subject. One type of question that Henderson's results posed was, for example, whether the lexical effects were upon the immediate processing of visual features or rather at a verbal recoding stage or even at a post-encoding stage of comparison or response decision. In the experiment reported next, Henderson & Chard (1976) sought the answers for the questions above.

6.2.2 - Henderson & Chard (1976)

Henderson (1974) found that meaning could produce a WSE, without the need for orthographical regularity. He posited that the WSE was due to phonological recoding for lexical access (name code) rather than due to a comparison between visual information corresponding to the strings. The idea that the use of a name code would be more economical and efficient was a plausible one, since real words have name codes that are highly practised and thus presumably more accessible than nonword codes. It has been a common finding in the literature that words are named significantly faster than pronounceable nonwords (Frederiksen & Kroll, 1976; Baron & Strawson, 1976; Foster & Chambers, 1973).

Henderson and Chard (1976) extended the work further still by a visual same-different comparison task experiment, this time by combining a WSE - which depended upon lexical access (Henderson, 1974) - with two manipulations which appeared to be purely visual. One consisted of stimulus degradation and the other was the manipulation of case (fbi vs. FBI). No effect was found with regards to stimuli degradation. However, they found that the WSE for acronyms was confined to the visually familiar case (e.g. RT(FBI) < RT(IBF) but RT(fbi) = RT(ibf)). This result was enough to refute any suggestion that phonological recoding mediated the effect, with an advantage for familiar strings of letter names. Clearly, it was visual familiarity which was involved and the visual code which gained lexical access for familiar items was case specific. The significance of this experiment resides in the fact that it was the first time that visual and phonological processes could be disentangled from each other.

6.2.3 - Seymour and Jack (1978)

Seymour & Jack (1978) adapted the same-different technique used by Henderson & Chard's (1976) and not only attempted to replicate their previous experimental results with upper-case acronyms (BBC), but also extended the investigation to examine well known abbreviations that are normally seen in lower-case (for example, *etc, yds, cwt*). For the upper-case versions, such as BBC, they replicated the results of previous experimenters. This added more support to the idea that at least for binary classification tasks, there is a direct visual access code to the mental lexicon. Again, the interaction found here between familiarity and case could not be explained in terms of the theories which assume phonological recoding to be an obligatory step preceding lexical access, since the upper- and lower-case versions of the abbreviations presumably possess identical phonological codes.

However, no evidence of interaction between case and familiarity was found for the lower-case familiar abbreviations. This same type of result has been found elsewhere (Baron, 1977) and Seymour and Jack put forward several reasons for the failure to find an interaction. First, lower-case abbreviations may be less frequently seen in isolation compared with running text. Second, the lower-case abbreviations formed a relatively heterogeneous group including initials of noun phrase (*sae, mph, lbw*) and contraction of words of varying syntactic class (*etc., neg., vol., yds.*). By contrast, the upper-case items used were a relatively more homogenous group, consisting of titles which appear lexically equivalent to proper nouns. These speculations lead to several testable predictions about the origins of the WSE. To test them what is required are stimuli that can be tested in both upper- and lower-case format. As will be seen later, this is exactly the type of stimuli that have been used in the new experimental evidence to be presented later in this chapter.

6.2.4 - Besner et. al. (1984)

The latter two experiments reported above prompted the notion that visual familiarity plays a role in recognition and that the visual code which gains lexical access for familiar items is case specific. It also brought the suggestion that acronyms are processed holistically. Recall, however, that the issue of whether words are processed analytically or holistically is an old (Cattell, 1886) as yet unresolved question.

What is not clear from the above results is whether the experimental findings are generalisable to all word types or whether they are specific to acronym processing. Instead of being processed alphabetically are they processed ideographically? Besner dismissed this idea in his doctoral thesis (1980) after having carried out some experiments with acronyms using the laterality framework. He based his experiments on evidence that was available at the time, that tachistoscopic report is superior with right visual field presentation (RVF) when the stimuli are printed alphabetically, but superior with left visual field presentation (LVF) when the stimulus is a single ideograph (Ellis & Young, 1977). This led to the proposal that logographs are better processed by the right hemisphere and alphabetic print better processed by the left. Thus, Besner reasoned, if it is true that letter identification is not involved in the processing of abbreviations, then an overall LVF advantage might be seen in tachistoscopic reports with acronyms. However, as pointed out above, when running experiments where he used acronyms as the stimuli material, he could find no evidence for a LVF advantage and concluded that acronyms were not being processed as logographs.

A dominant belief at that time was that cAsE aLtErNaTiOn in experimental stimuli did not have any impact on the WSE. By searching the WSE literature Besner and colleagues (1984) have shown that this was not the case and that the various experiments presented different patterns of results. However, they argued that the discrepancy among the results obtained by the various experiments, is due to huge variations in tasks used in the experiments. On the one hand, in all case alternation experiments that used the tachistoscopic technique, no alteration of the WSE could ever be found

(McClelland, 1976; Adams, 1979; Allport, 1979). On the other hand, experiments such as those reported by Pollatsek et.al. (1975) that had used the binary paradigm, did find that the magnitude of the WSE effect for "same" judgements was smaller when stimuli were case alternated than when they were not.

The observations above, together with the null results obtained with acronyms using the laterality framework, led them to the conclusion that it was not the case that the familiarity effects obtained in experiments using acronyms were due to any exceptionality of the stimuli. Instead, what seemed to be determining the presence of the familiarity effect was the task chosen by the experimenter. When the task required unique identification of the stimuli, as for example in tachistoscopic reports and naming, then no familiarity effect could be found. However, when the task could be accomplished by means of only some vague knowledge of the stimuli, for example, a matching same-different task or the performance of lexical decisions, then familiarity effects were present.

As a solution, they proposed two distinct types of processing in word recognition: the identification process and the figural familiarity process. The identification process uses information that is based only on the knowledge of the abstract identity of the individual letters composing a string. This information is used when the response uniquely specifies a stimulus, as for example, in tachistoscopic reports and naming tasks. The figural familiarity process, apart from using information at the individual letter level, also uses information at the word-level, for example, word-shape. It is applied to accomplish tasks such as the binary classification paradigm and lexical decision task (LDT) for which unique identification of the stimuli is not necessary.

In the wider frame of visual word recognition research it may be that the biggest impact of these studies have been on the area of modelling visual word recognition. For example, more recently the approach to modelling has been that of the development of *hybrid* models, that is, models designed to accommodate in themselves both types of routes: analytic and holistic ones. This echoes Besner and colleagues (1984) dichotomisation hypothesis, which suggested the split up of the visual word recognition process in at least two different types of processes: the identification (of analytic nature) and the figural familiarity (of holistic nature) processes. Some examples are Besner & Jonhston (1989) and Allen & Madden (1990) models of visual word recognition.

To sum up the four previously described experiments have established that visual familiarity is involved in word recognition and also that the visual code that gains lexical access for familiar items is case specific. Later, Besner and colleagues (1984) hypothesised the dichotomisation of visual word recognition in two different processes: the identification and the figural familiarity process. The identification process is the more relevant to the process of reading. The echoes of Besner and colleagues (1984) finds and proposal of dichotomisation are later found in the modelling visual word recognition.

6.3 - Motivation

We now come to the experimental investigation performed in the present study. Two types of experiments, namely, a LDT (Lexical Decision Task) and a naming task will be reported below. They were used to investigate if the hypothesis put forward by Besner and colleagues (1984) could be sustained when using brand names as the stimulus materials.

Brand names are a more appropriate class of words than acronyms to use as experimental stimuli for several reasons: First, as many brand names are part of our everyday life experience, they may become much more familiar to us than acronyms. This makes them an ideal set for testing environmental influences. Second, like acronyms, they are almost invariably represented using the same case. Third, they allow for further investigation into the effects of other visual features, such as the colour, size and font that are part of their configurational identity. Recall that no familiarity effects have been found for lower-case acronyms (Baron & Strawson, 1976; Seymour & Jack, 1978). It was speculated (see section 6.2.3 for details) that unlike acronyms, there are lower- and upper-case brand names belonging to same syntactic category. The reason for that, as mentioned before in section 6.2.3, might be due to the fact that, differently from the upper-case acronyms, the lower-case ones do not form a homogenous group in syntactic terms (e.g., *etc., sae...*). In the case of brand names, not only familiar upper-case but also familiar lower-case ones can be selected which belong to the same syntactic category. Most importantly, brand names have a more lexicalised pronunciation than acronyms, allowing for a more naturalistic type of experiment to be carried out involving naming. Fifth, they offer more choice, at the material selection stage, in terms of number of letters and number of syllables than acronyms.

6.4 - Experiments

6.4.1 - Experiment 4: Lexical decision task

6.4.1.1 - Participants

The participants were 28 volunteers from Heriot-Watt University in Edinburgh. All participants had normal or corrected-to-normal vision, and each was run individually in one session of about 20 minutes.

6.4.1.2 - Stimuli and design

The experimental stimuli were 48 brand names, 48 common words and 96 nonwords. The brand names were chosen from an initial list composed of 97 items, which are always written in capital letters. A group of 10 volunteers from the University of Edinburgh ranked this initial list for familiarity, using an ordinal scale from 1-7 points. The 48 brand names used are those whose median ranked over 5.5 points. The common English words used in the experiment were matched with the brand names so as to have the same number of letters and syllables. They are also high frequency words with more than 50 appearances in a million, taken from the CELEX database. The 96 nonword fillers were also controlled for number of letters and syllables. The stimuli were divided in 2 sets (A and B). In group A, half of the material was seen in entirely upper-case letters and the other half in lower-case. In group B the type-case was the reverse of group A. Half of the subjects saw one stimulus set, the remaining subjects saw the other set.

The set of brand names used in this experiment can be classified in two different groups: the mixed- and the pure-brands. The group referred to as the mixed-brand is composed of 12 items which are normally found in the mental lexicon due to the fact that they share names with common English words (e.g., SHARP). The other group, called **pure-brands** is formed by brand names whose names were invented for the purpose of labelling the products to which they refer (e.g., DULUX). Consequently, they do not appear in the mental lexicon functioning as common English words. The number of items composing the mixed-brand group amounted to twelve. Therefore, 12 matching items were selected from the other group. All the material in the two groups were carefully balanced with respect to their number of letters, number of syllables and rate of familiarity. See Table I in Appendix VI for how they were controlled.

6.4.1.3 - Procedure

A Macintosh computer running the PsyScope software (version 1.0.2) controlled stimulus presentation and timing. Participants were asked to sit in front of the screen and rest each forefinger on the leftmost and the rightmost buttons of a response box, located in front of them. The buttons were marked *yes* and *no*, respectively. Participants were instructed to decide as quickly and as accurately as possible, whether each of the strings appearing on the screen was a real word or not, by pressing the appropriate button on the response box. They were told that brand names would also count as real words. Each letter string stayed on the screen until the participant made a response. After each response a 2000 ms delay was introduced before the next string of letters appeared on the screen.

The participants were given six practice trials to familiarise them with the task. The experiment was divided in two parts, each lasting for about 5 minutes. Participants were allowed to rest between the two parts. The stimulus was displayed using New York font, 24 point and bold. Each of the subjects belonging to group A of the experimental design received a random sequence of strings. The randomisation of materials belonging to subjects in group B was matched to that of group A.

6.4.1.4 - Results

6.4.1.4.1 - Pure-brands x Mixed-brands

Initially an analysis was carried out to verify if the two different types of brand names, i.e., pure- x mixed-brands, would behave differently relative to capitalisation. The response time results for the two sets are presented in Table (6.1) below

	Pure-brands (e.g. DULUX)		Mixed-brands (e.g. SHARP)	
	RT(ms)	SD (ms)	RT (ms)	SD(ms)
lower-case	812.1	145.4	721.1	67.6
upper-case	757.3	126.2	678.7	51.0

Table (6.1) - Pure/Mixed-brands reaction times (ms) and standard deviations as a function of conditions per items.

An analysis of variance (ANOVA) was performed for subjects (F_1) items (F_2). A 2(letter-case) x 2(brand-type) repeated measures ANOVA was used for subject analysis. A main effect for brand-type was found F_1 (1,27) = 21.0, P < .001 and for capitalisation as well F_1 (1,27) = 12.7, p = .001. However, no interaction between the variables was found F_1 (1,27) = 0.15, P = .70. A mixed ANOVA was run for item analysis and a main effect was found for capitalisation was found F_2 (1,22) = 13.3, p < .001. such that capitalisation elicited faster RTs. However, no effect was found for brand-type F_2 (1,22) = 0.95, p = 0.3. Also, no statistically significant interaction was found between capitalisation and brand-type with F2 (1,22) = 0.44, p = .51. The main result here is the highly significance of the capitalisation variable.

Before going further, note that it is possible that the lack of a statistically significant effect for the brand-type variable in the F_2 analysis and also of an interaction between the type and capitalisation variables in both type of analysis, is due to the small sample of items used. A further larger experiment would be needed to settle the matter satisfactorily. However, this was not within the scope of this thesis. For the purpose of further analysis we collapsed all brand names into one group. Thus, from here onwards, no distinction will be drawn between pure- and mixed-brands. They are all grouped together and termed *brand names*.

6.4.1.4.2 - Brand names x Common English Words

The analysis that follows refers to the totality of material used in the experiment. For the complete list of items used, including common words and their familiarity rates, number of letters and number of syllables, consult the Appendix IV and V.

As it can be seen from the graph in Fig. (6.1), RTs were significantly faster for brand names in their familiar upper-case form when compared to the unfamiliar lower-case. No difference was found for common English words. An analysis of variance (ANOVA) was performed with both, participants (F_1) and items (F_2). A 2(letter-case) x 2(word-type) repeated measures ANOVA was used for the subject analysis. A main effect for capitalisation was found F_1(1,27) = 7.15, p < .01 and and also for the word-type variable F_1(1,24) = 56.1, p < .001. A 2(letter-case) x 2(word-type) mixed ANOVA was run for the F_2 analysis.

The word-type variable show a significant effect $F_2(1,94) = 28.8$, $p < .001$ and also the capitalisation variable $F2(1,27)= 7.15$, $p < .01$. RTs are significantly faster for brand names. Finally, an interaction between word-type and capitalisation was also found in both analysis $F_1(1,24) = 8.6$, $p < .007$ and $F_2(1,94) = 10.9$, $p < .001$, clearly showing that the facilitatory effect of upper-case format is holding only for brand names. These results will be discussed further in a later section.

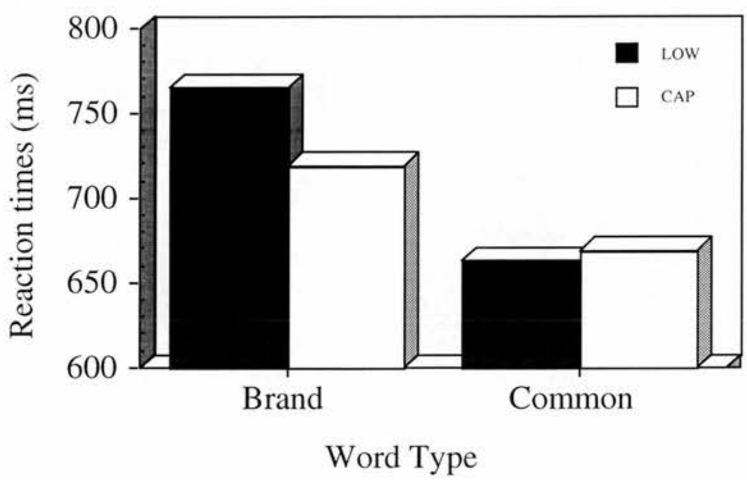

Fig.6.1 - Capitalisation effect on brand names x common words

The RTs corresponding to Fig. (6.1) and the percentage of errors can be seen in the Table (6.2) below. Note that the percentage of errors is very low. This reflects the generally high familiarity of both, the brand names and the words used. Note also that the RTs for brand names are longer than those for the words. Despite their familiarity, it can be suggested that brand names are not normally thought of having the same "word-status" as a word like *home* or *car*. This could lead to further processing, before the lexical decision is made, resulting in longer RTs.

	Brand names			English-Words		
	RT (ms)	SD (ms)	Errors %	RT (ms)	SD (ms)	Errors %
lower-case	765.1	132.1	0.9	682.5	116.8	0.3
upper-case	718.7	121.0	0.9	668.6	142.2	0.2

Table (6.2) - Average reaction times (ms), their standard deviations (SD) and Error rates a function of conditions for the capitalisation and word-type variables per subjects.

Further analysis was performed on the number of letters forming the brand-name strings. Brand names were grouped in two sets, the *short-group* composed of brand names 4-5 letters long and the *long-group* formed by brand names 6-9 letters long. To have the number of cells balanced one item was randomly chosen to be taken out, i.e., *yamaha*. See the table below, for items presented according to their number of letters. After grouping the brand names in the two sets, the average and SD reaction times for the items in each set were calculated. They are shown in Table V in Appendix VI.

An analysis of variance (ANOVA) was performed with subjects (F_1) and items (F_2). A 2 (letter-case) x (letter-number) repeated measures ANOVA was used. A main effect was found for the number of letters $F_1(1, 27) = 30.9$, $p < .001$. Similarly, a main effect was also found for letter-case $F_1 (1,27) = 35.3$, $p < .001$. There is no statistically significant interaction between letter-number and letter-case with $F_1(1,27) = 1.5$, $p = .24$. i.e., there is a word length effect in brand names.

A 2(length) x 2(letter-case) mixed-anova, was run for the item analysis. A main effect was found for the number of letters $F_2(1, 44) = 14.8$, $p < .001$. Similarly, a main effect was also found for letter-case $F_2 (1,44) = 12.8$, $p < .001$. There is no statistically significant interaction between letter-number and letter-case with $F_2(1,44) = 1.2$, $p = .27$. i.e., there is a word length effect in brand names.

	Short-brands		Long-brands	
	RT(ms)	SD (ms)	RT(ms)	SD(ms)
lower-case	721.1	85.6	852.7	140.4
upper-case	686.9	74.7	787.4	134.8

Table(6.3) - Short/Long brands reaction time (ms) and standard deviations as a function of conditions for the capitalisation and brand length variables per items.

The same type of analysis was carried out for common English words. See table II in Appendix VI where items are presented according to their number of letters. Two words were chosen at random from 6-9 set to be excluded during the statistical analysis.

Again, after grouping the words, the average RT and SD were calculated and are shown in Table (6.4) below.

	Short common-words		Long common-words	
	RT(ms)	SD	RT (ms)	SD
lower-case	668.7	60.9	649.9	64.1
upper-case	667.4	57.5	662.9	61.1

Table (6.4) - Short/Long common words reaction time (ms) and standard deviations as a function of conditions per items

An analysis of variance (ANOVA) was performed with subjects (F_1) and items (F_2). A 2 (letter-case) x (letter-number) repeated measures ANOVA was used for the subject analysis. No main effect was found for the number of letters $F_1(1, 27) = 0.29$, $p = .59$. Similarly, no effect was also found for letter-case $F_1(1,27) = 0.83$, $p = .36$. There is no statistically significant interaction between letter-number and letter-case with $F_1(1,27) = .42$, $p = .52$.

A 2(length) x 2(letter-case) mixed-anova, was run for the item analysis. No main effect was found for the number of letters $F_2(1, 42) = 0.57$, $p = .45$. Similarly, no effect was also found for letter-case $F_2(1,42) = 0.32$, $p = .57$. There is no statistically significant interaction between letter-number and letter-case with $F_2(1,42) = 0.48$, $p = .49$. This is an unusual result. We speculate that this might be due to the fact that because the words were all chosen to be of very high frequencies, we are witnessing a ceiling effect.

It was not possible to perform a analysis by syllable, as the vast majority of the material had only a short range of syllables, i.e., 2-3.

This lexical decision task produced a number of interesting results. First, the capitalisation variable in its two conditions produced statistically significant results for brand names but not for commons words. The same is true for the word-type variable. Analysis of member of letters were also performed for both types of words. We will return to these results at the end of this chapter in the discussion section. Comparisons with experimental results obtained with acronyms will be made. First, a second experiment performed using the same material has to be presented.

6.4.2 - Experiment 5: Naming

Having found a familiarity effect for brand names in the LDT, exactly as it would be predicted by Besner and colleagues' hypothesis, we proceeded to investigate the issue further, by running an experiment that used the same material of the previous one, however with the difference that the experimental paradigm used this time was the naming task.

6.4.2.1 - Participants

The participants were 40 undergraduate students from the Linguistics Department at Edinburgh University. They participated in partial fulfilment of course requirements. All participants had normal or corrected-to-normal vision, and each was run individually in one session of about 20 minutes.

6.4.2.2 - Stimuli and design

The experimental stimuli and design was the same used in the lexical decision experiment. However, because this is a naming experiment the nonword stimuli were excluded.

6.4.2.3 - Procedure

Again, a Macintosh computer running the PsyScope software (version 1.0.2) controlled stimulus presentation and timing. Participants were asked to sit in front of the screen holding a directional microphone in front of them and close to their mouths. They were instructed to name as quickly and as accurately as possible each of the strings that individually appeared on the screen. Each letter string stayed on the screen until the participant made a response. As in the previous experiment, after a participant response, a 200 ms delay was introduced, before the next string of letters appeared on the screen.

The participants were given a training set to allow them to familiarise themselves with the task. The experiment was divided in two parts, A and B, each lasting for about 5 minutes. Participants were told to have a rest interval if they needed so. The stimulus was displayed using font New York, size 24 points and bold characters. Each of the subjects belonging to group A of the experimental design received a random sequence of materials. The randomisation of materials belonging to subjects of group B was matched to that of subjects of group A.

6.4.2.4 - Results

No significant difference between upper- and lower-case was found for either type of word (brand name or common word). This is shown by the plot in Fig. (6.2) and is in contrast to the results obtained in the LDT experiment. Again, an analysis of variance (ANOVA) was performed with both participants and items. A 2(letter-case) x 2(word-type) repeated measures ANOVA was used.

The RTs for strings in the upper-case format were longer than for lower-case ones. However this trend was not significant : $F_1(1,39) = 2.2$, $p < .14$ and $F_2(1,94) = 2.73$, $p < .13$. Main effects were found for the word-type variable ($F_1(1,39) = 170.5$, $p < .001$ and $F_2(1,94) = 43.0$, $p < .001$) with longer RTs being found for brand names compared to common English words. Although not relevant to the investigation here, this result is only to be expected, since many of the familiar brand names used in the experiment are rooted in languages other than English and might have induced subjects to use non-native phonology. No interaction between word-type and letter-case was found: F_1 $(1,39) = 0.03$, $p< .86$ and F_2 $(1,94) = 0.14$, $p < .71$. This should make the task of naming them more difficult resulting in longer reaction times.

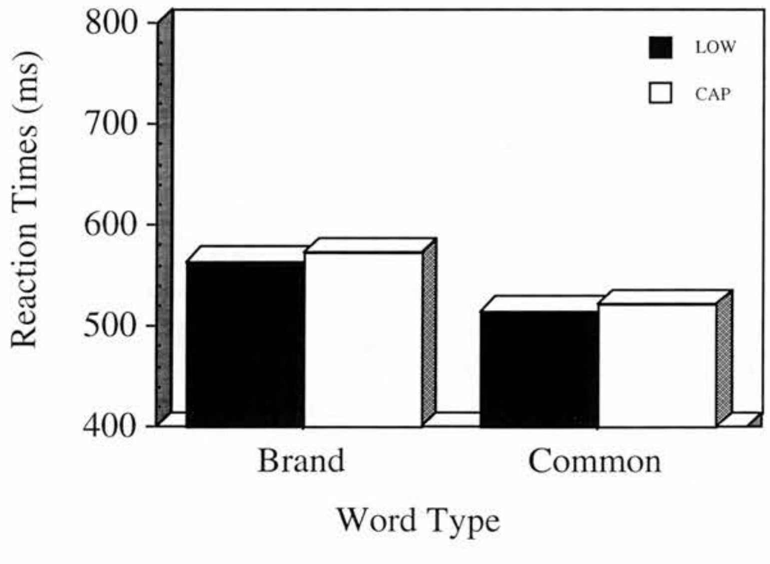

Fig. (6.2) - Lack of Capitalisation effect on brand names x common words

RTs and percentage of errors can be seen in the Table (6.5) below. Note that, as before, the percentage of errors is very low and again the brand names RTs are longer than those for common words.

	Brand names			English-Words		
	RT (ms)	SD (ms)	Errors %	RT (ms)	SD (ms)	Errors %
lower-case	563.7	66.7	0.1	514.7	60.2	0.07
upper-case	572.5	65.9	0.1	521.4	84.4	0.1

Table (6.5) - Average reaction times (ms), their standard deviations (SD) and Error rates as a function of conditions for the two variables(capitalisation and word-type) per subjects.

6.5 - Discussion

At the introduction we gave some of the reasons why brand names were chosen as stimuli in this study. One of them is that they have a consistent figural representation and some brand names have also become very familiar to us, since we are frequently exposed to them in our environment.

As seen before, in that experiment, RTs were faster for brand names presented in their familiar upper-case format when compared to the unfamiliar lower-case. As already discussed, the involvement of a visual code in the processing of written strings has already been acknowledged for some time now (Henderson, 1974; Henderson & Chard, 1976, Seymour & Jack, 1978). It has also been accepted that, at least in the case of acronyms, the visual code which gains access for familiar items is case specific (Henderson & Chard, 1976). Our findings in the LDT experiment not only support such views but also extend them to the brand names category. In previous experiments of this kind, psychophysical features of the stimuli (e.g., upper-case letters having a larger size than lower case) have been proposed as an alternative explanation for the results (Paap et. al, 1984). In the present investigation , the stimuli were carefully controlled, so that if these features came into play they could be readily detected. In particular, no statistically significant difference between RTs for upper- and lower-case common words was found. Therefore, any subtle difference between the letter-cases, by themselves, can not account for the results obtained for brand names. This allows us to rule out an explanation based only on psychophysical features in the present study. Instead, we suggest that these results show a clear manifestation of the familiarity effect: the letter-case (capital letters) seems to have become part of the identity of the brand names used in the study. This information seems to have been incorporated into the processing strategy of the participants. This seems to improve their performance in the LDT, resulting in shorter reaction times. It should also be pointed out that this explanation is valid for the case of acronyms as well. It is plausible to believe that information about letter-case is an inherent part of the identity of acronyms such as FBI, BBC, etc.

The results reported previously concerning the further analysis of word-length effects, that is the search for influences connected with the total number of letters composing the strings are quite interesting. No effect was found for common words. However, brand names presented a strong length effect in the direction of shorter strings showing faster RTs than the longer ones. Furthermore, although the interaction between the case and length variables did not reach statistical significance a trend in the direction of short upper-case strings being the fastest was found.

There has been some disagreement on the topic of word length in the literature, concerning the LDT. Frederiksen & Kroll (1976), for example, did not find any array-size effect in a LDT and this is in agreement with our findings for the common English words used. However, Balota & Chumbley (1984) reported relatively large length effects for words when using the "same" task. They reasoned that a lack of control in that earlier experiment with regards to the length and regularity of the items used could have propitiated some sort of contamination by other sources that prevented the manifestation of the array-effect. Their explanation can also account for the results obtained here and I suggest that the source of contamination in the present study is the very high frequency of the common English words used in this experiment. In this case, the word frequencies are so high that their familiarity becomes the dominant factor, overwhelming completely the possible word length effect.

Now, before discussing the results obtained for brand names in the LDT experiment, let us consider the lexical status of brand names. There is no work done on this subject to the best of my knowledge. However, it is plausible to suggest that people are not used to think of them as simply being common English words. Supporting this idea, there are the results obtained here, where RTs for brand names were always longer than those for common English words (the material was all controlled for array length and number of syllables). Although this result was obtained for both tasks (LDT and naming), it is particularly relevant in the case of the LDT where the items were controlled in all possible respects. Unfortunately, with the naming task, due to the difficulties in finding items that fit all the necessary criteria it was not possible to have control of all variables, for example, having brand names and common words controlled for initial articulation. Thus, when asked in LDT to press the *yes* button for brand names it took participants slightly longer to acknowledge the brands as being real words.

Another possible explanation for brand names presenting longer latencies than common words is that of a frequency effect, i.e., reaction times to brand names took longer due to the fact that brand names would have lower frequency countings than the common words used. However, notice that, some of the brand names (e.g., TESCO) used are very common indeed, i.e., they are seen in written form, much more frequently than, for example, the word *table*. Unfortunately, there are no frequency databases around that include brand names. One way to disentangle the issue, is to run further experiments where equal familiarity ratings are to be used for both words and brand-names.

Earlier I have suggested that it was the very high frequency nature of the common English words used in the experiment that masked the effects of length. Now, I suggest further that the array effect found for brand names could have been enhanced by this brief hesitation in acknowledging brand names as real words. It might be said that it had exactly the opposite effect of that caused by high frequency, causing brand names to show high statistically significant and strong length effect (131.1 ms of difference for lower-case and 100.5 ms for upper-case).

In admitting that frequency could be playing a role in the effects discussed above, an issue could be raised that, frequency instead of familiarity might be the variable responsible for the differences found in terms of capitalisation between brand names and common words. However, it should be noticed here that to discredit the familiarity explanation common words would have to present a capitalisation effect in the opposite direction to that of brand names, since we have seen in our earlier experiments and literature has shown as well, that latencies should be faster for lower-case strings compared to upper-case. Although not significant if one look at the figures for common words, this is the direction to be found there. In the case of brand names however, the effect is exactly in the opposite direction what could only be explained in terms of familiarity.

As for the second experiment, using the naming paradigm, the absence of a familiarity effect for both categories, i.e., brand names and common words is in sharp contrast with the results discussed above for the LDT. However, they are not different from those reported in the literature (for a different class of words), when similar paradigms were used, as for example, tachistoscopic reports (Besner, 1980). In fact, the pattern of results obtained here matches well previous findings. This point takes us to the discussion of what has been our main purpose for running the present experiments. This was to test the hypothesis put forward by Besner et. al. (1984) concerning the dichotomy in visual word recognition of the processes of figural familiarity and identification. We can confirm that this dichotomy is present for the category of brand names. It must be added that the brand names showed this effect in a straightforward manner. Previous experiments using acronyms had inherent difficulties, as acronyms (e.g., FBI) are not normally pronounced the same way as common words. So, a tachistoscopic report had to be used, instead of the more straightforward naming task. Brand names, having a more lexicalised pronunciation, seemed therefore the ideal candidates for exploring these questions further. On the one hand, a familiarity effect was found for the LDT, that supposedly does not demand any specific knowledge about the identity of the string to be processed. On the other hand, no effect could be found for brand names during the naming task, which is supposed to demand unique identification of the string before it can be processed. These results are in keeping with Besner and colleagues hypothesis, thus favouring the analytic models of visual word recognition.

6.6 - Conclusions

To conclude, brand names are part of our everyday life and as such should be of interest to all those modelling visual word recognition either naturally or artificially. Besides, they also offer a fresh opportunity for looking again into many of the issues that in earlier studies became restricted due to limitations of using only acronyms as stimuli. For example, they are a better source than acronyms due to their rich figural representation to run lateralisation experiments which investigate the differences between the logographic and normal processing of strings. It is interesting, to notice that in our experiments, even after having been stripped of all its graphical representations, still an effect of capitalisation was found for brand names. Also, the difference between mixed and pure brand names if taken further can throw further light into the organisation of the lexicon. In neuropsychological investigations the category of brand names has been neglected. However, it can be intuitively accepted that it is probably one of most familiar category of words that can be found in the lexicon. An exciting finding about brand names is that apart from the situations mentioned here, there are quite a number of other situations that can also be investigated using them as stimuli materials. The most important contribution of this chapter, is the discovery of the category of brand names as a source of material for further psychological investigation. Concerning the mechanisms of processing in visual word recognition, our findings here are in the direction of Besner and colleagues' hypothesis that poses that more than one type of processing strategy is used in visual word recognition. Furthermore, these strategies appear to be task dependent. However, as it is normally the case in science further investigation is required if any final conclusion is to be reached on the subject.

Chapter 7

Grapheme-phoneme probabilities in British English

The importance of pronunciation studies for the understanding of the mechanisms of language processing has already been discussed in this thesis. As an estimate of the complexity of the task for the speaker of assembling a pronunciation from a written form it is useful to look at the nature of the mappings from letter to sound.

The work reported in this chapter has been born out of the large number of different pronunciations that were found for the nonwords that were used in Experiment 2 (Chapter 4). The prospect of being able, at the end of the work, to compare the probabilities with which subjects output a pronunciation with the probabilities given by a large database of the English language was exciting enough to make us invest our time in this work. Thus, here we report a computational assessment of the mappings between orthography and phonology in English. The study comprises an exhaustive search of the CELEX database and an analysis of the frequencies of occurrence of grapheme-phoneme correspondences in British English. This information is used to estimate the pronunciation of any string of English graphemes[1]. In particular, we predicted the pronunciation of real English words and compared this with the pronunciation given to them by the CELEX database. The usefulness and limitations of this pronunciation prediction algorithm are discussed. The algorithm was further assessed by examining its behavior for

[1] The term "grapheme" will be used here to indicate a letter or group of letters that corresponds to a single phoneme (e.g., Venezky, 1970).

nonwords. A corpus of nonword transcriptions was collected from a set of trained phoneticians: these provided the baseline for assessment of the algorithm.

7.1 - Motivation

In this section we discuss in a general manner four areas of studies where the type of statistical knowledge described in this work can be useful. The first of these is perhaps one of the most applied branches of language research, i.e., the work being developed in speech synthesis and recognition. In the 1960s it was expected that within ten years computers would be interacting intelligently with humans through speech. Despite the disappointment produced by such excessively optimistic hopes, progress has been made in this field, for example, in the area of telephone services. However, there are still a host of problems that have to be overcome so that more sophisticated speech systems can be developed.

Most of the speech recognition systems nowadays are designed based on a statistical approach to identifying words. This is computationally less expensive than using neural networks or artificial intelligence techniques. Such systems use models of phonemes and identifying these has proved to be faster than using whole words. The first part of the work described in this chapter has been the creation of an exhaustive database of the frequencies of the grapheme-phoneme correspondences in English. In the speech recognition context this type of information can be very useful, for example, in the development of dictation systems. To translate the spoken words being dictated into words on the computer screen an accurate mapping between phonemes and graphemes has to be built.

A second area where this framework can bear fruits is that of spelling research. Nowadays, one of the motivations behind the study of spelling processes is, as noted by Brown and Ellis (1994), the increasing realization that the processes of learning to read and learning to spell are intimately related. Frith (1986) and Ellis (1994), for example, argue that the developmental process of learning to read can not be fully understood without researchers paying attention to the concomitant processes of spelling development. Through repeated practice in spelling, the child may come to appreciate the subtle relationships between a symbol in the written word and its corresponding sound in the context of the spoken word. The discovery of this relationship is the key to alphabetic insight (Liberman & Shankweiler, 1979).

A less theoretical and more practical approach to the subject is that of the teaching of spelling. Educators regard children learning to spell correctly with the utmost seriousness, since they argue that, perhaps more than being linguistically important, to spell correctly is socially important (Czerniewska, 1992). Spelling is often used as the criterion for distinguishing the educated from the uneducated. Therefore, there is a lot of debate and research devoted to finding the best methods for teaching spelling. The most traditional method of teaching the spelling of English is also very controversial, i.e., the use of lists of isolated words. These lists usually comprise everyday words which are graded according to their difficulty of spelling. As an example, let us take the work published by Shonell and Brown (1995) entitled "A spelling list for seniors" that is composed of over 3000 words. As the authors explain in the introduction of their work the lists were based on pragmatic principles alone, i.e., they were based on a survey of spelling difficulties among pupils in secondary schools. Although part of the traditional methods of teaching spelling, these tables are frequently criticized. One of the oldest criticisms is that it is impossible to predict which words a child will write during his school life, and it is even more clearly impossible to predict which words he will need at any particular stage in his school career or at any stage in his verbal development (Bennett, 1967). I would like to pose a further criticism with regards to the creation of these tables: they are not based on a real understanding of the cognitive processes commanding spelling. This is perhaps more pertinent to those interested in the more theoretical approach to the subject, i.e., to unveil the cognitive processes of spelling than in the immediate results it offers.

Psycholinguists are becoming more and more aware of the wealth of information that cross-linguistic studies can bring into their fields of studies. A third area where the type of statistics we have used here can be applied is in connection with the orthographic depth hypothesis. It states that the extent to which readers use phonological information in recognizing words depends on the extent to which it is represented by the orthography of their language (Seidenberg, 1995). We refer the reader to section 4.2 (Chapter 4) where this subject has already being discussed. We suggest that the creation of a method of classification by which the various languages could be ranked in terms of how deep or shallow their orthography is, could be very useful. For example, such a ranking system could help in the design of more efficient cross-linguistic experiments in the area of word recognition. The differences between languages such as English (deep orthography) and say, Portuguese and Spanish (shallow orthography) are not difficult to grasp. However, differences become less obvious between two languages closely related such as Portuguese and Spanish mentioned above. Therefore, the development of computational programs to extract the relationships between phonology and orthography in a language could produce an objective measurement of the level of orthographic depth. We will come back to this subject again later in the chapter.

The deep orthography of English allows for the existence of two kinds of words - regular words, which follow the rules of spelling-to-sound correspondence and exception words, which break them. There is a large number of psycholinguistic studies dedicated to the investigation of the different types of behaviour that occur in relation to these two kinds of words. This interest was partly induced by several findings in the literature that regular words take less time to pronounce than exception words (e.g., Glushko,1969; Gough & Cosky, 1977). The type of algorithm being described here can also be useful to those pycholinguists interested in gathering experimental material to test regularity effects in the English lexicon. For example, by extracting the probabilities with which the regular word *gave* would be pronounced as /geIv/ and the exception word *have* would be pronounced as /həv/.

Finally, as it has been mentioned earlier in this chapter our main inspiration for computing the probabilities of grapheme-phoneme correspondences was to compare the probabilities between human performance in the pronouncing nonwords and the probabilities contained in a large database of the language referring to that pronunciation. In Section 7.7 the results obtained with an experiment performed with a group of phoneticians will be compared with a corpus of nonword transcriptions.

7.1.2 - Background

The present study is based on Berndt, Reggia & Mitchum (1987) work on pronunciation probability. They used a corpus of 17,310 words (Hanna, Hanna, Hodges & Rudorf, 1966) to provide probability estimates for the pronunciation of particular graphemes in English. Probabilities were derived for the correspondences between individual graphemes (i.e., letters or letter clusters corresponding to a single phoneme) and their phonemic realizations. As will be seen shortly, the present work uses a corpus a thousand times larger, which will produce more representative results.

An important aspect of the Berndt et. al. work and also of ours, is that the calculated probabilities are context independent; i.e., they do not systematically reflect word-specific morphological, syllabic and suprasegmental information. As seen earlier, these aspects can be expressed as rules that are used to map strings of graphemes to sound. Thus, the probabilities provided here are a conservative estimate of the extent to which particular letters and letter clusters are pronounced as particular phonemes in English. They do not provide any information about the rules responsible for the derivation of these correspondences. As this work demonstrates, a purely statistical knowledge of the language is not sufficient to capture many of the subtleties of pronunciation. Nevertheless, it shows in what situations this statistical information can be used to provide useful estimates of pronunciation, especially when nonwords are involved.

7.2 - The statistical calculations

All the statistical information on word frequency used in this study was taken from the CELEX database. The frequency information given in that database in its turn, was derived from the COBUILD corpus of the University of Birmingham. In the 1991 version extracted and corrected by CELEX, this vast database contained approximately 17.9 million words, taken from written sources of many kinds. Frequencies of occurrence for each word of the corpus are included. The final form of the CELEX database is in a table format, with each word and its associated information appearing as a row, divided into many columns. The first column contains a number uniquely identifying a wordform, and the second column has the wordform itself. A third column gives the frequency of occurrence of that wordform in the COBUILD corpus and various other columns provide a wealth of information. Of relevance here is a column containing the phonetic transcription of the wordform. Therefore, our aim here is to extract from this database, the phonetic transcription corresponding to each grapheme present in the English language. This was done by taking one grapheme at a time and determining if it is present in each wordform (each row) of the CELEX database. When it is found in a wordform, its corresponding phoneme must be searched in the phonetic transcription column. The frequency for the grapheme, pronounced as the particular phoneme found is then retrieved from the third column of the database. This basic search algorithm poses many challenges and subtle nuances, which will be described below. It is important to emphasize the large size of the corpus from which the CELEX database was derived. As mentioned earlier, the previous attempt at deriving grapheme-phoneme correspondences used a corpus of only 17000 words. The fact that we use a one thousand times larger corpus has two important and related consequences. Firstly, it gives a better statistical estimation of the frequencies of graphemes which appear less frequently in the corpus. Also, many new grapheme-phoneme correspondences not obtained before were detected and their probabilities of occurrence calculated. Note also that in the previous study the grapheme-phoneme probabilities could not be derived in a straightforward manner. The corpus had been created from phoneme/grapheme associations, i.e., starting from spoken words; their graphemic counterparts had been obtained. Through a computational process, the reverse probabilities, that is, the desired grapheme-phoneme correspondences were then calculated. In the present work, there was no need for this further complication process, as the Grapheme-phoneme correspondences are already available in the CELEX database, albeit in the form of words and their phonetic transcriptions. Finally, the phonetic transcriptions in the CELEX database are based on British English. Thus, a new kind of Grapheme-phoneme correspondence is derived, as the previous study was based on American English.

7.3 - Preliminary statistics

Before delving into the computational aspects of the grapheme-phoneme searches, some preliminary and simple statistics might be useful to give an idea of the database used. The total number of lines (number of wordforms) in the database is 160595.

DISK	IPA	DISK	IPA	DISK	IPA	DISK	IPA	DISK	IPA	DISK	IPA
p	p	r	r	x	X	I	I	u	u:	c	æ
b	b	f	f	h	h	E	ɛ	3	ɜ:	q	ɑ̃:
t	t	v	v	w	w	{	æ	1	eI	0	æ̃:
d	d	T	θ	J	tʃ	V	ʌ	2	aI	~	ɒ̃:
k	k	D	ð	_	dʒ	Q	ɒ	4	ɔI		
g	g	s	s	C	ŋ̩	U	ʊ	5	əʊ		
N	ŋ	z	z	F	m̩	@	ə	6	aʊ		
m	m	S	ʃ	H	n̩	i	i:	7	Iə		
n	n	Z	ʒ	P	l̩	#	ɑ:	8	ɛə		
l	l	j	j	R	*	$	ɔ:	9	ʊə		

Table (7.1) I - Correspondence between the CELEX (Disk)[2] and IPA character sets. See text below.

The number of characters in each word and the number of phonemes in each word transcription were counted. CELEX has 3 different sets of characters for phonetic transcription. In this chapter, the CELEX (Disk) representation of

[2] Four different sets of phonetic character codes are available from CELEX. They are SAM-PA, CELEX , CPA and DISK. The three first use ASCII codes to represent certain of the IPA characters. The DISK transcription was chosen here for technical reasons due to its unique single character nature.

phonemes will be used throughout. The IPA (International Phonetic Alphabet) transcription of phonemes will be given as well, in parentheses. In the "Disk" character set, each phoneme is represented by a single ASCII character, apart from "x", which is transcribed as "ks", "gz" and "kS" (/ks/, /gz/ and /kʃ/) and one of the possible pronunciations of the grapheme "u", when transcribed as "ju" (/ju:/). This almost one to one correspondence between the phonemes and the characters representing them greatly simplifies the task of searching and counting the phoneme occurrences. Table (7.1) shows the correspondence between the CELEX (Disk) characters and the IPA character set.

The average length and standard deviation of word-types and their phonetic transcriptions are (8.3±2.7) characters and (6.9±2.4) characters, respectively. Next, the number of words containing one, two, three, and so on, phonetic syllables, was computed. It can be seen in Table (7.2) below that the vast majority of words have two, three, one or four syllables, in this order. There is only one "word" with 12 syllables: "European economic community".

Syllables	Words
1	24913
2	61737
3	45652
4	20210
5	6245
6	1433
7	314
8	58
9	22
10	7
11	3
12	1

Table (7.2) - Distribution of words with different number of syllables in CELEX.

This serves to show that CELEX lists not only isolated single words, but also some expressions, phrasal verbs, abbreviations, etc. Still, the vast majority of entries in the database consist of isolated single words. In order to keep all the simple numerical results in one place, we will advance two small results from the present study. The total number of unique graphemes searched was 195 and the number of Grapheme-phoneme correspondences was 464. Some of the preliminary results presented above will be used later on, when deciding on the best strategy to search graphemes embedded in "difficult" words.

7.4 - Grapheme search strategies

Various programs were written in the AWK language to search for the frequencies of occurrence of the graphemes. The graphemes and various other variables containing regular expressions were passed to the programs as command line arguments. The Grapheme-phoneme associations from Berndt et. al. were used as a starting point to build our own database of associations. On finding a new phoneme corresponding to one of the graphemes of Berndt et. al. but not present there, it was added to the database. Also, various new graphemes and their phonetic transcriptions were detected and included in the database resulting from this work.

7.4.1 - Searching bigrams, trigrams and longer graphemes

Long graphemes are easier to search as they generally occur only once in a word. Therefore, prior to counting the frequency of occurrence for one of these graphemes, a simple program was run to check if it appeared more than once in any of the words of the database. For example, only three trigrams were found to repeat in words: "igh", "sch" and "tch". All words found for each grapheme and their frequencies were stored in a separate file, to be further processed later. With this provision, all trigrams and longer graphemes can be considered to appear only once in each word. This greatly simplifies the searching algorithm. To give an example of how the "repeated" graphemes above are treated, take "igh" for example. It appears twice in 15 words, all variations of the root "highlight". The sum of occurrences of these words is 155. Therefore, after searching for the "igh" grapheme, taken to appear only once in each word, another 155 occurrences have to be added to the number of occurrences found. Obviously, this simple manual procedure becomes impracticable for graphemes which repeat in many words, because of the size of the temporary files created to store them.

The search strategy used to count the frequency of trigrams and longer graphemes is now described briefly. The search for a grapheme-phoneme correspondence starts by passing the grapheme and its transcriptions to be searched, to the program. A real example might be of use here. Take for example the grapheme "igh". It was searched using the following command:

```
nawk -f big grapheme="igh" regexp="[^ae]igh|^igh" transcriptions= "2" nossofile
```

"nawk" in the expression above stands for "new awk", the programming language used. The parameter "-f" means that the program itself is stored in a file, whose name comes next, i.e., "big". The next two expressions give the grapheme "igh" and a 'regular expression' variable regexp="[^ae]igh|^igh". This simply means that instances of "igh" to be counted must not be preceded by "a" or "e", because in these cases, they form other graphemes, namely "aigh" and "eigh". Because of the way nawk functions, it will search for "igh" always preceded by some character, if only the first part of the regular expression ([^ae]igh) had been present. The rest of the regular expression (| ^ igh) ensures that the grapheme "igh" can also appear at the beginning of words. The next command line argument lists all the transcriptions to be searched, in this case, only one "2" (/aI/). The last argument "nossofile" gives the file containing the CELEX words, their frequencies and phonetic transcriptions.

A detailed analysis of the programming code is not appropriate here however, a practical example will be given, to illustrate the procedure. When words containing the grapheme "igh" above are identified by the program, a trial search "*width*" is computed. The search *width* is simply given by the difference between the number of characters in the word and in its phonetic transcription, not counting diacritics. Take for example the word "enlightenment' and its transcription "InI2tHm@nt" (/InlaItn̩mənt/), which have 13 and 10 characters respectively. In this case the search *width* is therefore 3. For many words, these two lengths are identical, in which case the search *width* is set to 1. Now the position k where the grapheme starts in the word is computed and the transcription is searched for, at positions k ± *width* in the phonetic transcription. In the example above, k = 4 and *width* = 3. Thus, the phoneme will be searched for, in the positions from 1 to 7 in the phonetic transcription and one transcription (the character "2" (/aI/), as passed by the command line argument) is found. Therefore, in this case the search for "igh" pronounced as "2" (/aI/) succeeds without further complications. When k = 1 or k is the last position in the phonetic transcription, then it is not possible to use the *width* as calculated above. In these cases, the phoneme is searched at positions from k to k±2×*width* respectively. According to the database, the word "enlightenment" appears 45 times in the corpus. Therefore, this number is added to the sum of occurrences of the grapheme "igh" pronounced as "2" (/aI/) and a new database line is read. If the grapheme "igh" occurs in it as well, the search algorithm above is repeated, if not the next line is read.

More complex situations can occur and two of them will be discussed next. Consider for example the grapheme "mb", which is pronounced as /m/ in words like "climb", "plumber", etc. There are however words containing "mb", such as "numbers" or "chambers", in which "m" and "b" are both pronounced, i.e., they are two separate graphemes. Words such as these are excluded from the frequency count of the grapheme "mb".

A more difficult problem occurs with graphemes which have many different pronunciations, such as "ai", which can be pronounced in 7 different ways (see the Appendix VII): "2", "8", "@", "E", "{", "I" and "1" (/aI/, /ɛə/, /ə/, /ɛ/, /æ/, /I/ and /eI/). Next, take the example word "traitors" and its transcription "tr1t@z" (/treItəz/). The word has 8 characters, whereas the phonetic transcription has 6, giving a search *width* of 2. The bigram 'ai' starts at position 3 in the word, so the starting position for the search is 3, with *width* 2. It is obvious from the transcription that in this case two valid pronunciations for the bigram "ai" are found: "1" and "@" (/eI/ and /ə/). We know however that the bigram should appear only once in each word. In the example above, the "@" (/ə/) found actually corresponds to another grapheme namely, "or". The solution to this problem is simple. Every time the number of pronunciations found is larger than one, the search *width* is reduced by one and the search is repeated. If two or more pronunciations are still found, the *width* is reduced by one again and a new search is performed. This procedure is repeated until only one pronunciation is found. In the example above, when the search *width* is reduced from 2 to 1, a single pronunciation (the correct one, as the other corresponds to the grapheme "or") is found. This word occurs 50 times in the corpus and this number is added to the sum of the grapheme "ai" pronounced as "1" (/eI/).

There are words however, where the *width* is reduced to one and still, two pronunciations are found. When the *width* is further reduced to zero, no pronunciation results. One example of this situation for the "ai" bigram above is the word "assailing" and its transcription "@s1lIN" (/əseIlIŋ/). Note that the word has 9 characters, whereas the phonetic transcription has 6, giving a search *width* of 3. The bigram starts at position 4 in the word, so, the starting position for the search is 4, with *width* 3, i.e., the search goes from position 1 to the end of the phonetic transcription. In this case, three possible pronunciations, "@", "1" and "I" (/ə/, /eI/ and /I/), will be found. The *width* is then reduced successively to 2 and 1 and still two pronunciations will be found "1" and "I" (/eI/ and /I/). When it is reduced to zero, because the starting position of the search is the fourth character ('l') in the transcription, no pronunciation is found. When this situation occurs (search *width* is zero, but no pronunciation is found), the starting position of the search is shifted by one to the left, i.e., position 3 in our example. The *width*, given by the difference between the lengths of the word and its transcription, is computed again and the whole procedure is repeated. It can be seen in our example that in this case the *width* will still be reduced to zero, but because the search now starts at position 3, the correct pronunciation corresponding to the grapheme is found.

The process of shifting the start of the search by one to the left is arbitrary and might assign the wrong (in the sense that it corresponds to a different grapheme, with the same pronunciation) phonetic transcription to some graphemes. It can be justified though, on empirical terms, as follows. It was shown in the preliminary statistics that, on average, the words are longer than their phonetic transcriptions by 1.4 characters. Thus, the phoneme corresponding to a given grapheme is, on average, to the left of the position occupied by the grapheme in the word.

Under certain circumstances, the grapheme being searched is identified in a word, but for some reason its corresponding phoneme is not found in the transcription. In this case, the whole line containing the word, its frequency of occurrence and transcription is printed on the screen and also sent to a file, for further manual processing. This situation can occur for many reasons, the most common being the case where a new phonetic transcription (not included in Berndt et.al.) is associated with the grapheme under search. This was the case with the grapheme "ai" used in the examples above. In Berndt et. al, there are only 5 phonemes associated with it: "1", "E", "I", "@" and "{" (/eI/, /ɛ/, /I/, /ə/ and /æ/). The present algorithm found that in the CELEX database it has two more possible pronunciations: "2" and "8" (/aI/ and /ɛə/). The program was then run a second time, including these new pronunciations in the "transcriptions=" command-line argument. A few obviously wrong transcriptions in the CELEX database were also uncovered by this method, which forced problematic words to be printed on the screen. Another case that produced this output was the single letter grapheme "h", but this was a genuine case in which the grapheme has no corresponding phoneme. Single letter graphemes will be discussed in the next section below.

RESULTS OF CELEX SEARCHING grapheme: ai transcriptions: 128EI@{ regular expr: ai[^gn]|ai$		
SUM OF OCCURENCIES OF TRANSCRIPTIONS:		
grapheme	transcriptions	occurencies
ai	2	359
ai	8	26266
ai	@	3394
ai	E	49817
ai	{	178
ai	I	8484
ai	1	44838

Table (7.3) - Search results for grapheme "ai".

When the program reaches the end of the CELEX database, all occurrences of the given grapheme/phonemes associations have been counted. The program then produces a table containing in each line, the grapheme, one of its associated phonemes and the number of occurrences found, as exemplified by Table (7.3), for the grapheme "ai". This table is saved to a file (without the headings, i.e., only the lines containing a grapheme, a phoneme and its number of occurrences) and subsequent runs of the program, for new graphemes, appends their tables to this file. In this way, a new database, containing Grapheme-phoneme associations and their frequency of occurrences was generated.

7.4.2 - Searching single letter graphemes

The search strategies used in this case are similar to the ones described above. This time however, the grapheme can repeat any number of times in a word. This is a further complication that has to be dealt with in the search strategy. Once a word is found to contain the grapheme under search, a check is made to determine how many times it appears in the word. If it occurs only once, the search proceeds as in the previous section. If it appears n times ($n \geq 2$), all positions it occupies in the word are stored. A search is then triggered, starting at the first position of the grapheme, with the search *width* given as before, by the difference between the word length and the length of the phonetic transcription. After finding the transcription corresponding to the first occurrence of the grapheme, another search is started (from the second position found) to find the second phoneme, corresponding to the second occurrence of the grapheme. This process is repeated n times, until all occurrences of the grapheme are accounted for.

Again, all problems with the search *width* and the 'shift left' procedure described above come into play. In a word such as "however", transcribed as h6Ev@R (/haʊɛvə*/), the search position for the first "e" is 4, which corresponds to "v" in the transcription. It is clear that there are two possible phoneme associations, one character away from the starting position, "E" and "@" (/ɛ/ and /ə/), on each side of "v". During the search, the *width* will be reduced to zero and the starting position will be shifted by one character to the left, before a single phoneme corresponding to this starting position can be found. The word "however" appears 7516 times in the corpus. Thus, this value is added to the occurrences of "e" pronounced as "E" (/ɛ/) and also to "e" pronounced as "@" (/ə/). This way multiple occurrences of a grapheme in a word are taken into account, even when different instances of a grapheme in a word are associated with different phonemes, as above.

In the process of searching the database, many new grapheme-phoneme associations, not present in Berndt et.al., came to light. All such cases have been included in the database storing the results of these searches. In all, 464 Grapheme-phoneme associations were found, with 195 unique graphemes, giving 2.4 phonemes associated with each grapheme, on average.

7.5 - Computing the probabilities

Having obtained the Grapheme-phoneme associations and the number of occurrences for each of them, it is now possible to calculate the percentage of occurrences of each grapheme in the corpus. First, all occurrences for all graphemes were added, producing a grand total of 69181671. Dividing the total number of occurrences of a grapheme in the corpus by the grand total above (×100%), gives the percentage for occurrence of the grapheme in the corpus. This ratio of the grapheme total and the grand total can also be viewed as a good estimate for the prior probability of occurrence of the given grapheme in the (British) English language. If the grapheme "a" is taken as an example, its total number of occurrences is obtained by adding the number of occurrences for all its associated phonemes. This gives a total of 3746713 and the ratio of this number and the grand total is 0.054. Therefore the prior probability of appearance of the grapheme "a" (in all its phonemic associations) in the corpus is 5.4%. It should be noted that this is a reasonably high percentage, because it does not include combinations of "a" and other characters, as these are other independent, graphemes. For example, when "a" is followed by "i", a new grapheme, i.e. "ai", is formed. The "a" figuring in the grapheme "ai" was not considered in the counting above.

As the number of occurrences of each grapheme-phoneme association has also been computed, the percentages of occurrences for each individual grapheme-phoneme association can be obtained as well. For example, "a" pronounced as "@" (/ə/) occurred 591123 times in the CELEX corpus. Dividing this by the total number of occurrences of grapheme "a" (3746713) gives the percentage of occurrences of "a" being pronounced as "@" (/ə/). The number obtained is 0.158, which can be interpreted as the probability (×100%) of "a" being pronounced as "@" (/ə/). The probabilities for the remaining phonemes associated with "a" have been calculated in the same manner. Similar calculations were performed for all graphemes and for all grapheme-phoneme associations, producing the corresponding grapheme and grapheme-phoneme probabilities. In the Appendix VII, all grapheme-phoneme associations are listed, together with both probabilities, for the grapheme as a whole (prior probability) and for each of its phonemic associations. Also included in the Appendix VII is an example word, its phonetic transcription and its number of occurrences in the CELEX database, for each grapheme-phoneme association.

7.6 - Database validation and consistency checks

After obtaining all probabilities, simple tests were made, to ensure consistency and to check the database for any mistakes. It is clear from the discussion above, that the grapheme prior probabilities are normalized to 1, i.e., the sum of all grapheme prior probabilities P_{gi}^{p} adds to 1:

$$\sum_{i} P_{gi}^{p} = 1 \quad (1)$$

where P_{gi}^{p} is the total (prior) probability of appearance of grapheme g in the database.

The grapheme probabilities P_{gi}, taken for each grapheme separately, are also normalised to 1:

$$\sum_{i=1}^{k} P_{gi} = 1 \quad (2)$$

where k = number of grapheme-phoneme associations for grapheme g. Short programs were written to perform these sums in the database and both types of probabilities were found to be correctly normalized.

In order to validate the statistical information contained in the Appendix VII, a simple procedure was adopted. Perhaps the best possible test for the consistency of the statistical database compiled in this work is to use it to predict the pronunciation of real words taken from the CELEX database. This will check the accuracy of the statistical information obtained, against some of the words (and their phonetic transcriptions) used for its derivation. Strictly speaking, actual pronunciation can not be predicted solely on the basis of the grapheme-phoneme associations in the Appendix VII, as stress information is not available there. Still, it should serve to retrieve the correct phonemes contained in a given word.

A number of different schemes can be employed to use the statistical information obtained before, to predict word pronunciation. Metrics based on the present statistics, coupled with rules, perhaps similar to those devised by Venezky(1970), could prove very effective in pronunciation prediction. However, only the simple metric proposed in Berndt, et. al. was employed to test the database statistical content and to provide a simple means of comparison with previous work. Let P(p|g) be the probability of grapheme "g" being pronounced as phoneme "p" and P_g^* be the probability of the most frequent correspondence for the grapheme "g". For instance, $P_A^* = 0.496$ and P(ə|A) = 0.1578, P(æ|A) = 0.496 and so on (see Appendix VII). Consider a printed word to be given by a sequence of graphemes (containing one or more characters each) $g_1g_2,\ldots,g_n$. Its correct pronunciation would then be given by a string of phonemes $p_1p_2,\ldots,p_n$. An overall metric of predictability of the pronunciation of such a word is given by Berndt, et.al.:

$$m = \frac{1}{n}\sum_{i=1}^{n}\frac{P(p_i \mid g_i)}{P_{gi}^*} \tag{3}$$

Clearly, $P(p_i|g_i)/P_{gi}^*$ is simply the ratio of the probability of grapheme g_i being pronounced as p_i and the most probable pronunciation of grapheme g_i, P_{gi}^*. Therefore, eq. (3) above is simply the average of these ratios, for all the *n* graphemes in the word. This metric of predictability was used in Berndt, et. al. to calculate the probability with which the statistical grapheme-phoneme correspondences would predict correctly the pronunciation of a given word. In the present work we go one step further and use Eq. (3) to obtain the probabilities of all the possible pronunciations of a given word. For example, it can be seen in the Appendix VII that there are 3, 9 and 1 phonemes associated with the graphemes "t", "a" and "p". Therefore, the number of possible pronunciations of the word "tap" will be 3×9×1 = 27. Let t_i be the number of phonemes p_i associated with grapheme g_i. In general, the number of possible pronunciations $p_1p_2,...,p_n$ of a word given by the sequence of graphemes $g_1g_2,...,g_n$ will be $t_1 \times t_2 \times ... \times t_n$. Obviously, the correct pronunciation will not always be the one having the highest (m=1) probability. Using the same examples given in Berndt, et.al. for comparison, our database produces for the word "TAP" :

T → t [P(t|T) = 0.940 $P_T^* = 0.940$]
A → æ [P(æ|A) = 0.496 $P_A^* = 0.496$]
P → p [P(p|P) = 1 $P_P^* = 1$]

Therefore:

$$m = \frac{1}{3}\left(\frac{0.940}{0.940} + \frac{0.496}{0.496} + \frac{1}{1}\right) = 1$$

Thus, the statistical information included in the Appendix VII produces the correct pronunciation, in this case. However, as seen above, the word "TAP" has 27 possible pronunciations associated with it. Despite this relatively large number of possible pronunciations, the correct pronunciation is the one producing the highest value of m. In conclusion, it can be said that the pronunciation of the word "TAP: is highly predictable from the grapheme-phoneme associations of its component graphemes.

Taking the other example (the word WHO), in Berndt, et. al.:

WH → h [P(h|WH) = 0.204 $P_{WH}^* = 0.796$]
O → u [P(u|O) = 0.162 $P_O^* = 0.363$]
So,

$$m = \frac{1}{3}\left(\frac{0.204}{0.796} + \frac{0.162}{0.363}\right) = 0.35$$

Two things should be noticed in this case. Firstly, the pronunciation with the highest m (m=1) is "wQ" (/wɒ/), but this is not the correct pronunciation. There are in this case 18 (9×2) pronunciations and if they are sorted in decreasing order of m, the correct pronunciation "hu" (/hu:/) occupies the 12th position. The second thing to notice is that m = 0.35 for the correct pronunciation, as opposed to m = 0.09, given in Berndt, et. al. This discrepancy can be explained by the vast size of the corpus on which the present statistical information is based. This should be more representative of the frequency of occurrences of graphemes in (British) English and produced a higher probability of "O" being pronounced as "u" and also of "WH" being pronounced as "h". Therefore, even if the correct pronunciation is not the most probable, it could be argued that this value of m is more accurate than that found in Berndt, et. al. Especially for low-frequency grapheme-phoneme correspondences, the size of the corpus used can have a significant effect on the probabilities obtained. The fact that American and British English were used in the previous and in the present work respectively, will undoubtedly change the probabilities of certain grapheme-phoneme correspondences.

It would be extremely tedious to compute manually all the possible pronunciations of any given word and their associated probabilities. This is obviously a task best performed by a computer program. Accordingly, a program was developed which, given a word already divided in graphemes, searches the statistical database and computes all possible pronunciations of the word and their associated probabilities. Before displaying the results, they are sorted in decreasing order of probability and the number of possible pronunciations to be displayed can also be controlled by the user. The task of dividing a given word in its associated graphemes has not been tackled in this work and has to be performed by the user.

Using this program, many interesting aspects of the quality of the statistical information contained in the Appendix VII can be probed. It was found that certain words, despite being long and containing many vowels, have highly predictable pronunciation. One such example is the word 'understand', which has 9 graphemes and, because they have many different pronunciations, produce a large number of possible combinations. The total number of possible pronunciations for this word is 272160 ($=10\times3\times2\times7\times4\times3\times9\times3\times2$) and it is remarkable that the program finds the correct pronunciation as being the most probable, based solely on the simple metric of predictability in eq. (3), i.e., m = 1 for Vn-d@-st{nd (/ʌn-də-stænd/). On the other hand, there are deceptively simple words, such as "examine", for which the statistical information alone can not be used even as a first approximation to the correct pronunciation. This word has 7 graphemes and 2592 possible pronunciations; the correct one is given as the 805th by the program. Even in this case though, the failure of the metric used is not as bad as it might appear. The value of m obtained in this case was m = 0.83, which is still very close to the most probable, m = 1. In this case the program fails to detect the correct pronunciation with any accuracy, but it still gives important information about the word. There is a large number of possible phoneme combinations (pronunciations), all having very close values of m, between m = 1 and m = 0.83 (for the correct pronunciation). This type of information can be useful to guide the choice of material for experiments on phonology, for example.

The program above was also used to test the probabilities of the grapheme-phoneme correspondences. This was done by using the program to predict the pronunciation of words taken from the CELEX database. The results produced could then be compared with the actual pronunciations also listed in CELEX. Obviously, the frequency of occurrence of a word will have a big influence on the predictability of its pronunciation. In a high frequency word, at least some of its graphemes will also have a high frequency of occurrence. Thus, the correct pronunciation of high frequency words should be easier to predict. Five sets of a hundred words each, chosen at random, were extracted from CELEX for the tests. The sets 1, 2, ..., 5 contain words whose frequencies of occurrence are larger than or equal to 0, 100, 1000, 5000 and 10000 respectively. Therefore, words in set 1 are chosen completely at random from the CELEX database. The others are still chosen at random, but with the provision that their frequencies are above the threshold set for each group.

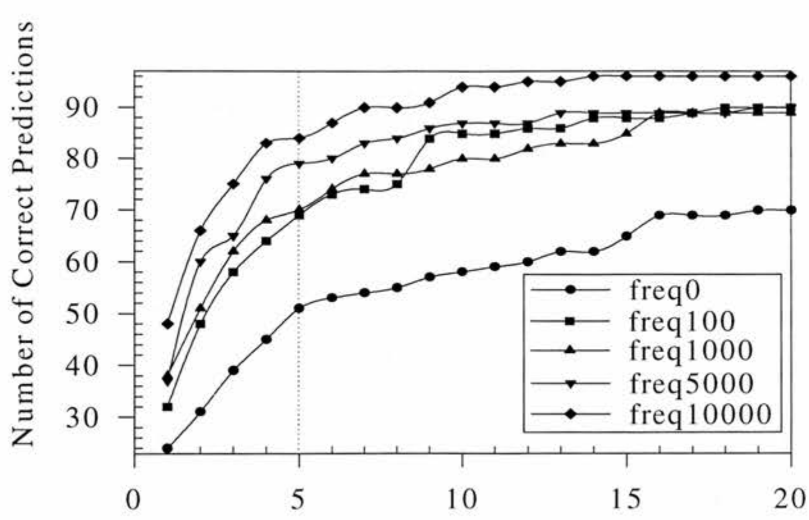

Fig.(7.1) - Predicting the pronunciation of words chosen at random from CELEX. The higher the frequency of a word, the easier it is to predict its pronunciation

The program was run for each individual word in each group and the position of the correct pronunciation (i.e., if it was the 1st, 2nd, 3rd, 4th, and so on, in decreasing order of m) was saved. Fig. (1) shows the results of this test graphically, where only the first 20 most probable pronunciations were taken into consideration. Thus, if the pronunciation produced by the program is at a position beyond the 20th, it is considered that the program failed to detect it. For each set, the ordinate axis represents the number of words whose pronunciations can be predicted with the first, first two, ..., first twenty pronunciations. Therefore, if the program is "allowed only one guess", it can predict the pronunciation of 24, 32, ..., 48 words for the sets 1, 2, ..., 5, respectively. This general trend of higher predictability with higher frequency of occurrence of the words is maintained throughout the sets. Thus, this test can be considered as a validation of the grapheme-phoneme correspondences extracted from CELEX.

During the testing described above, it was noticed that the shorter the word, the easier it was in general, to predict its pronunciation. This is not surprising, as the majority of words in the corpus from which the probability database was derived have fewer than 4 phonetic syllables (see section 7.3).

This motivated one further test, where only monosyllables were used. Again, five sets of a hundred monosyllables each, with frequencies larger than 0, 100, 1000, 5000 and 10000 were chosen at random from the CELEX database. For the 5000 and 10000 sets, there were only 79 and 63 monosyllables respectively in the corpus. So, the results for these two sets were normalized to 100 words. As before, the program was run for each individual word and the results are shown in Fig. (7.2), where the axes have the same meaning as in Fig(7.1). The first thing to notice is that the five curves are much closer together here than in Fig. (7.1), i.e., the frequency of the monosyllables in each set is less important than it is for words of uncontrolled length (Fig. 7.1).

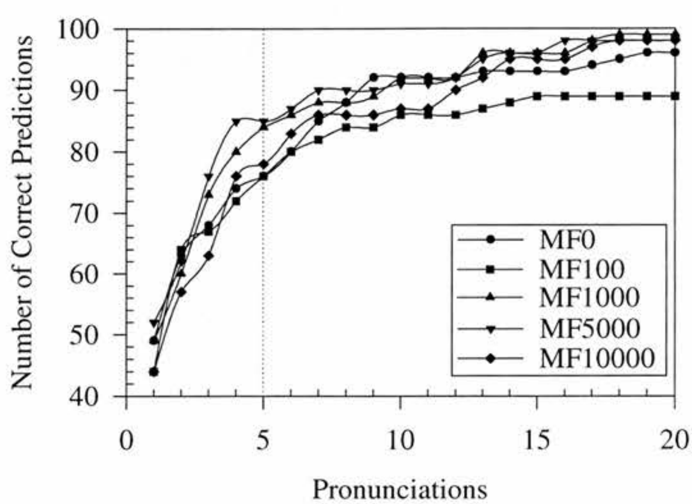

Fig. (7.2) - Predicting the pronunciation of monosyllables chosen at random from CELEX. The frequency of occurrence has less effect here than in Fig. (7.1).

As expected, the monosyllables have pronunciations easier to predict, compared with the words used in Fig.(1). If a level of 5 pronunciations (5 allowed guesses) is chosen, then the program can predict the pronunciation of 51, 69, 70, 79 and 84 words in the five sets of words used in Fig. (7.1). For the monosyllables (Fig. (7.2), the corresponding figures are 76, 76, 84, 85 and 78. Because there are not enough monosyllables with frequencies above 10000, the statistics for this set of words is not as reliable as for the others. This could explain the fact that less words were correctly predicted for this set in five guesses than for the 1000 and 5000 sets.

The extensive testing presented in Figs. (7.1) and (7.2) demonstrate that probability alone is not enough to produce high levels of accuracy in pronunciation prediction. However, if the goal is accurate pronunciation prediction, than the statistical information obtained can be combined with rules, perhaps similar to those of Venezky, to produce a better metric for measuring the pronunciation predictability. On the other hand, the type of statistical information presented here is interesting in itself and might be useful to control the "frequency variable" in experiments involving the pronunciation of strings of graphemes not found in the (British) English lexicon.

Thus, the probability database in the Appendix VII might be used for other purposes as we have pointed out earlier. It can easily calculate the "phonetic neighborhood' of a given word, for example. In this context, it is interesting to recall the two example words discussed before, "understand" and "examine". When applied by those using nonwords and pseudowords in pronunciation experiments, grapheme-phoneme statistics can become an important tool for the selection of materials. In the next section, it is used to predict the pronunciation of nonwords created by the algorithm discussed in chapter 3. Pronunciations produced by a group of professional phoneticians were then compared with the results produced by the program.

7.7 - Nonword pronunciation prediction

One of the most prominent roles played by nonwords in psycholinguistic studies was that of becoming a test case in the debates of models of word recognition with regards to the role of rules in explaining behavior, and the adequacy of the parallel distributed processing approach. The Seidenberg and McClelland (1989) connectionist model of pronunciation has been harshly criticized for not maintaining its level of performance regarding the pronunciation of nonwords at the same level as that of human beings. However, this situation changed drastically when Plaut et. al. (1994), by improving the input and output representations of their model, were able to simulate successfully the reading of nonwords at the human level performance. See sections 4.3.3.1 and 4.3.3.2 in Chapter 4 for more a detailed account of these models and their results with regards to the simulation of nonword reading.

The set of nonwords used to test these models were those used in Besner & McCannan (1987) and Glushko (1979). Compared to the nonwords we have used in Experiment 1 (Chapter 4) the nonwords mentioned above can be regarded as "tamed" in their range of possible pronunciation. The nonwords used in Experiment 1 were created so as to have more than one pronunciation. This was largely achieved by the device of having them constructed so that their nuclei support different pronunciations (e.g., *fier*, *seid*). Also, both high- and low-frequency onsets and codas were used in the construction of the nonwords. As explained earlier, this method of creating them produced items that were orthotactically unusual (e.g., *coew*) alongside more usual ones (e.g., *hean*). Note that although in Experiment 1 subjects complained of the difficulty in pronouncing some of the nonwords, they managed the task well and all of them without exception could produce a pronunciation for each of the items presented. What we suggest here is that a corpus of nonwords such as the one mentioned above sets a new challenge to the rival models of pronunciation.

Next we describe an experiment where a subset of these nonwords has been used. Our subjects in the experiment were trained phoneticians and all of them were also native speakers of English. The choice of phoneticians as subjects was motivated by two reasons. First, the task required a certain amount of skill in phonetically transcribing English. Second, their transcriptions would certainly reflect their knowledge of Standard Southern British English. Both of these requirements were important because the aim of our experiment was to check the probabilities of these transcriptions against the database of grapheme-phoneme representations we have built. Recall that our database is based on the transcriptions in the CELEX database which are designed to reflect the Standard Southern British accent.

7.7.1 - Experiment 6

7.7.1.2 - Participants

The participants were 6 volunteers from the University of Edinburgh. All participants were phoneticians and native speakers of English.

7.7.1.3 - Stimuli and design

The experimental stimuli were a subset of 20 nonwords from those used in Experiment 1 (Chapter 4). Half of the material was *high frequency* nonwords and the other half low-frequency nonwords ("high-high" and "low-low").

7.7.1.4 - Procedure

Subjects were given the list of nonwords on paper. One word was written per line. They were asked to assign to the nonwords as many phonetic transcriptions as they thought possible. However, they were advised that when transcribing the nonwords they should keep in mind that the nonwords were based in the English language. There were no time restrictions. They were told to return the list when they had finished.

7.7.1.5 - Results

The results were organised as follows: for each of the nonwords, all transcriptions produced were collected and tabulated, counting the number of phoneticians who produced them. For example the nonword "seid" was transcribed as /saId/, /sid/, /sɛd/ and /seId/, by 3, 1, 3, 3 phoneticians respectively. In addition, the program described in sections 7.5 and 7.6, was used to calculate an m-score for each of the phoneticians' transcription. These are displayed on Table (7.4).

Table(7.4) - Nonwords and their transcriptions produced by Phoneticians

NP = number of phoneticians producing a particular transcription.
m-score = Probability of the corresponding transcription produced by the program. Where the m-score is marked with "-", the program failed to produce that particular pronunciation.

Nonwords	Transcriptions	NP	m-score
chreolc	/kril/	1	-
	/krilk/	1	0.82
	/kriəlk/	2	0.59
	/kriɒlk/	1	0.70
	/tʃrIəlk/	1	0.83
	/kriɔ lk/	1	0.59
shrooln	/ʃrun/	2	-
	/ʃruln/	5	0.98
	/ʃrulən/	1	-
	/ʃrəuɒln/	1	-
ptaamf	/tamf/	2	0.79
	/pətamf/	1	-
	/tæmf/	2	1.00
	/təmf/	1	0.83
sneertz	/snirts/	3	0.84
	/snirtz/	1	1.00
	/sniətz/	2	0.47
	/sniəts/	1	0.70
gryk	/grik/	2	0.75
	/grIk/	5	1.00
	/graIk/	1	0.81
	/grʊk/	1	-
thout	/θaʊt/	6	0.71
	/θɔt/	1	0.44
	/ðaʊt/	1	1.00
hean	/hin/	5	1.00

	/hɛn/	1	0.82
baif	/beIf/	5	0.98
	/bɛf/	2	1.00
	/baIf/	2	0.67
zielts	/zilts/	4	0.68
	/ziəlts/	1	0.69
	/zj ɛlts/	1	0.67
knyl	/nIl/	4	1.00
	/naIl/	1	0.74
	/knIl/	3	-
	/naIəl/	1	-
	/knul/	1	-
toond	/tund/	6	1.00
	/təʊɒnd/	1	0.75
seid	/saId/	3	0.72
	/sid/	1	0.82
	/sɛd/	3	0.67
	/seId/	3	0.68
mees	/mis/	4	1.00
	/miz/	5	0.92
spleij	/splaI/	1	-
	/splɛIʤ/	1	1.00
	/spleIʤ/	1	0.81
	/splei/	2	-
	/spleʤ/	1	-
	/sple/	1	-
gnuentz	/nunts/	1	0.75
	/nuənts/	1	-
	/nents/	1	-
	/nuɛnts/	1	0.41
	/gnuɛnts/	1	0.52
	/njuənts/	1	-
	/gnuntz/	1	0.79
	/gnɛnts/	1	-
fier	/fir/	3	0.78
	/faIr/	2	0.78
	/faiə/	1	-
	/fajr/	1	-

	/fiə/	3	-
jits	/ʤits/	1	0.75
	/ʤIts/	5	1.00
psuavs	/suəvz/	2	0.94
	/psuəvz/	2	0.95
	/swavz/	1	
	/psavz/	1	-
nueld	/nuld/	2	0.92
	/nɛld/	1	-
	/nild/	1	-
	/nuɛld/	3	0.81
	/nuəld/	1	-
	/njuld/	2	-
	/nwɛld/	1	-
sprouv	/spruv/	2	0.92
	/sprəʊ v/	2	0.80
	/spraʊ v/	3	1.00
	/sprʊv/	1	0.89
duast	/duəst/	1	-
	/duast/	2	1.00
	/djuəst/	2	0.39
	/dwast/	2	-
rhadz	/ræʤ/	3	1.00
	/raʤ/	3	0.79
	/rəʤ/	1	0.83
froitsch	/frɔItʃ/	6	0.80
ghuisch	/guʃ/	2	0.75
	/guIʃ/	2	0.77
	/guItʃ/	1	-
	/guiʃ/	1	0.51
	/gIʃ/	1	1
	/gɔIʃ/	1	-
	/gwIʃ/	1	0.75

Table (7.4) - Nonwords and their transcriptions produced by Phoneticians.

A Pearson's r coefficient correlation was calculated to establish the relationship between the number of phoneticians producing one pronunciation and its corresponding m-score. A positive correlation was found ($r = 0.44$, $d.f. = 84$, $p < .001$). The result shows that more phoneticians tend to produce a transcription that bear a high m-score. Seen from a different angle, this result is also a demonstration of the credibility of m-scores as a quantitative measure of pronunciation probability.

Next, a paired t-test was run to test for any differences among m-scores regarding the two different types of nonwords used in the experiment, i.e., high-high x low-low frequency. These can be seen in Table (7.5). The average m-scores were calculated in the following way: For each nonword, the original m-score given in Table (7.4) was taken and multiplied by the number of times that a particular transcription was given. This was done for all the transcriptions for which an m-score could be found. Those transcriptions without an m-score were discarded. The numbers obtained were then added and the result was divided by the total number of transcription produced for that nonword. For example, the calculation of the average m-score for the nonword "baif" are:

/beIf/	5 x	0.98	= 4.9
/bɛf/	2 x	1.00	= 2.0
/baIf/	2 x	0.67	= 1.3
Average m-score			= 8.2 / 9 = **0.91**

The average m-scores and number of pronunciations are given in the Table (7.5):

High Frequency	Number of Pronunciations	M. score	Low Frequency	Number of Pronunciations	M. score
grik	4	0.91	chreolc	6	0.96
thout	3	0.71	shrooln	4	0.98
hean	2	0.97	ptaamf	4	0.88
baif	3	0.96	sneertz	4	0.74
knyl	5	0.95	zieltz	3	0.68
toond	2	0.97	spleij	6	0.90
seid	4	0.64	gnuentz	8	0.62
mees	2	0.96	psuavs	4	0.94
fier	5	0.78	sprouv	4	0.91
jits	2	0.96	rhadz	3	0.88
nueld	7	0.85	ghuisch	7	0.66
duast	4	0.69	froitsch	1	0.80

Table (7.5) -The number of m-scores and number of pronunciations for low-low x high-high frequency nonwords

A Man-Whitney analysis showed that there is no difference between the levels of probability between the "low-low" and "high-high" frequency nonwords (t = 129, U=31, p =0.2).

A second Man-Whitney analysis was run for the number of *different* pronunciations. Again, no significant results were obtained (t=166, U=31, p = .32) and both results are discussed below.

The number of missed m-scores (i.e., not found by the program) was normalised in terms of percentage countings for both high-high and low-low frequency nonwords. A t-test was run to verify if any difference could be found between them. Again it was found that there are no statistically significant differences between the two groups (t= 158, U = 31, p = .73). The normalized m-scores can be seen in Table (7.6).

Low-frequency	No. of missing m-scores (%)	High-frequency	No. of missing m-scores (%)
chreolc	17	gryk	25
shrooln	75	thout	0
ptaamf	25	hean	0
sneertz	0	baif	0
zieltz	0	knyl	60
spleij	67	toond	0
gnuentz	50	seid	0
psuavs	50	mees	0
sprouv	0	fier	60
rhadz	0	jits	0
froitsch	0	nueld	71
ghuisch	43	duast	50

Table (7.6) - Normalized number of missing m-scores for high-high and low-low frequency nonwords.

In English, the vowels have a large number of possible pronunciations. This leads to many possible alternative pronunciations for a string containing vowels and to more variability in such cases. Therefore, it might be expected that the program will fail to compute m-scores for certain pronunciations, due to variant transcriptions of the nuclei. This was the case as shown in table (7.7).

onset	nucleus	coda
3	22	6

Table (7.7) - Number of missing m-scores in terms of onset/nucleus/coda.

7.8 - Discussion

As mentioned before, although pure statistical knowledge based only on one-to-one grapheme-phoneme correspondences can not be seen as the whole explanation for pronunciation, it still has a highly influential role in the process. Accordingly, we predicted that the nonword pronunciations produced by the phoneticians would have high m-scores. When the actual m-scores were calculated by the program for each pronunciation, this was found to be the case. Furthermore, a statistically significant positive correlation was found between the number of times a transcription was given and its m-score.

The majority of m-scores obtained were of high probability, with only few of them being rated under 48%. This finding could be interpreted as indicating that the same statistical mechanisms used to pronounce real words seem also to be involved in pronouncing nonwords. The frequency of pronunciation of the parts forming a monosyllabic nonword (onset, nucleus, coda) seem to determine to a great extent the pronunciation of the nonword as a whole.

When comparing the list of transcriptions produced by the program with that produced by the phoneticians, an interesting pattern of results emerges. The program was written so that if a string has 3 graphemes, for example, its phonetic transcription will also be composed of three phonemes. However, the nonword segmentation strategies used by the phoneticians, in some cases, do not seem to obey the same strict rules of the program. This is easy to see in the case of the nonword "knil". According to the statistics (and also the rules proposed by Venezky), the English grapheme "kn" should be transcribed as /n/. However, when transcribing the nonwords 3 and 1 phoneticians respectively produced the "alternative" transcriptions /knIl/ and /knul/. In these cases, the string was segmented with k and n being regarded as independent graphemes. Another example, this time however involving not the onset part but the coda is the nonword "spleij" and some of its transcriptions. Two of its possible transcriptions were /spleI/ and /splaI/. Here the grapheme "j" at the end of the string was omitted by the phoneticians, producing two pronunciations which can not be reproduced by the program.

We suggest that every transcription for which an m-score can not be derived can be seen as evidence of use of alternative phonologies. This behavior might be associated with the fact that the stimuli used are not real words but nonwords that resemble in many cases foreign words. The possible pronunciations could come in this case from a larger pool of phonologies than that of pure English. This result is interesting considering that our professional phoneticians were warned in advance that all the nonwords in the experiment were based on the English language. We find support for such interpretation in Brennen's "set size plausible phonologies" theory, described earlier in section 4.1.2.3 (Chapter 4). Recall that according to his theory, a less restricted phonological domain would be required for the pronunciation of nonwords compared to that of real words.

Another example of the interference of different phonologies affecting the pronunciation of unfamiliar words come from the studies reported by Fit (1997a, b). She investigated the spelling of unfamiliar names by asking Scottish subjects to write down and also to repeat aloud following a spoken prompt sixty town names from six different countries. She found that although subjects were not wholly accurate in their pronunciations, they did not always pronounced the names using English rules, even for languages they were unfamiliar with. They produced some non-English segments and consonant-clusters, and used non- grapheme-to-phoneme correspondences appropriately. However, in other situations they overgeneralised to languages in which the rules do not apply, suggesting that the native language is not always used for pronouncing nonwords.

One of the main themes that has being discussed in this thesis is that of the idiosyncrasies of proper names pronunciation. I have suggested that for connectionist models to become more psychologically acceptable they must incorporate means of reproducing the same type of behaviour we have seen here. To this effect the type of nonwords that have been tested could be an excellent means of investigating the networks capabilities.

As it can be seen from Table (7.7), the largest number of missing scores was derived from deviations in the pronunciation of the nucleus (22), followed by the coda (6) and finally onset (3). This result agrees with those obtained by Treiman et. al. (1996) where, by investigating lexical statistics it was shown that vowels have a wider variety of pronunciations than consonants. In that study it was also shown that orthographic and phonological units larger than single graphemes and single phonemes play a role in the description of English spelling-sound relations and in their use by fluent readers. For example, links between individual graphemes or between a vowel and final consonant in a word, help readers to deal with the eccentricities of English.

Further, we speculate here that the reason for failing to find any statistically significant results regarding the difference between the high-high and the low-low nonwords is due to the insufficient number of stimuli used (10 in each category) and the restricted number of subjects (6 phoneticians). Therefore, a larger study should be conducted if any conclusions are to be drawn from these comparisons.

As mentioned earlier, with the help of the programs developed here, it is possible to quantify the orthographic depth of different languages. Recall that the orthographic depth of a language is defined in terms of the extent with which the phonological information of a language is captured by its orthography. Thus, the orthographic depth can be calculated by taking the total number of grapheme-phoneme correspondences of that language (GPC) and dividing it by the total number(g) of graphemes in the language. Therefore, we are postulating that the ratio (GPC/g) is a direct measurement of orthographic depth of a language. In the case of English for example, we have seen in section 7.3, that the total number of graphemes (g) found in the CELEX database was 195 and the total number of grapheme-phoneme correspondences (GPC) amounted to 464. Thus, the orthographic depth of English is calculated to be GPC/g = 2.4.

A list of Portuguese grapheme-phoneme correspondences can be found in the Collins Gem dictionary (Harland, 1987), which contains the 57 most frequent graphemes. It is remarkable to find that the number of grapheme-phoneme correspondences for this (incomplete) list is only 61. This gives Portuguese an orthographic depth value of 1.07 (much lower than English).

In summary, for British English GPC/g = 2.4 and for Brazilian Portuguese PGC/g = 1.07, which shows, as would be expected by a native speaker, that Brazilian Portuguese has a much shallower orthography than British English.

Another possible way of measuring the orthographic depth of a language can also be derived from the present work. In order to implement it, a large number of high frequency words have to be randomly selected. Their possible pronunciations and m-scores are then calculated by the program described above. The words are then checked to determine if the program correctly predicted their pronunciation, i.e., if $m= 1$ for the correct pronunciation. The number of words for which this is true is divided by the total number of words searched and multiplied by 100, to give a percentage. It is this percentage which gives another measure of the orthographic depth of the language. For example, in section 7.6, (Fig. (7.1)), this percentage is found to be 48% for the group of words with frequency above 10 000. As mentioned before, this is actually a low value, which is now attributed to the deep orthography exhibited by British English.

At present, we have no means of computing these percentages for other languages. However, we predict that this percentage would be much higher for Brazilian Portuguese.

Finally, it can be seen that the two methods of measuring orthographic depth proposed here are inter-related. In a hypothetical language where each grapheme has only one corresponding phoneme, the first method would produce $GPC/g = 1$. This is actually the number to be obtained for the language with the shallowest possible orthography. At the same time, there would be only one possible pronunciation for each word in such a language (see eq. (3) and the discussion following it, in section 7.6). Therefore, the m-scores would always be equal to 1 and the resulting percentage of words whose pronunciations can be predicted correctly would be 100%. Again this would be a measure of very shallow orthography, according to our second method of depth measurement.

On the other hand, for a hypothetical language where each grapheme corresponds to, say 10 possible phonemes, the $GPC/g = 10$. This would be a language of very deep orthography. In addition, even for short words, composed, for example, of three graphemes only, there would be 1000 possible pronunciations. It would be unlikely that the pronunciation with $m= 1$ would be the correct one for a large number of words. Therefore, the percentage of words whose pronunciation could be predicted correctly would be small.

Again, we would conclude that such a language would have a deep orthography.

In conclusion, the statistical approach to language studies can be a very useful one. Here we have seen for example, how it can aid areas of studies such as speech recognition and synthesis and also spelling studies. Also, we saw how the calculation of the probability of pronunciation of unfamiliar strings can give us insights into the psychological processes that underlie pronunciation. We have shown that even in the case of professionally trained phoneticians, it is not possible to predict all the pronunciations of English nonwords that they produce. This is a reflection of the different string segmentation strategies used by them (and by other speakers), when recognising a nonword. Finally, we proposed two different (but related) methods of measuring the orthographic depth of a language. In this context, it was found that British English has a much deeper orthography than Brazilian Portuguese.

Chapter 8

Conclusions

The work reported in this thesis examined two different aspects of familiarity processes involved in visual word recognition. The first is how capitalisation influences visual word recognition. The second is the role played by onset, nucleus and coda in nonword recognition. These issues have important implications, both for theories of word recognition and also for its connectionist implementations. In what follows I will begin by summarizing the results before drawing more general conclusions and highlighting possibilities for further research.

8.1 - Summary of Thesis findings

8.1.1 - The role of initial capitalisation

One familiar visual aspect of English is that proper names are most of the time represented with an initial capital letter. We have seen, that there are many examples in the psycholinguistic literature that suggest that proper names are a special category of words (Cohen & Burk, 1993; Reason & Lucas, 1984; Cohen & Faulkner, 1986). To investigate any initial capitalisation influence in word recognition the following question was asked: Does initial capitalisation play any role in the processing of unfamiliar strings of letters? Two experiments using nonwords which manipulated initial capitalisation were run and they show that subjects produce fewer pronunciations for initially capitalised nonwords than for non-capitalised nonwords. As a result, I suggest that in English initial capitalisation does influence recognition by acting as a cue that is strong enough to prompt readers to perceive an unfamiliar string of letters as belonging to the category of proper names. I argue further that, as a consequence of this, a more restricted phonological domain is used to retrieve the pronunciation for initially capitalised strings in comparison to that used to retrieve the pronunciation for unfamiliar strings which are not initially capitalised. This is exactly the type of result predicted by Brennen's theory of plausible phonologies, which states that proper names possess a more restricted phonological space than that of nonwords. A possible mechanism that could account for our findings here is suggested in terms of Harm and Seidenberg's (in press) connectionist model of pronunciation and acquisition of reading skills.

8.1.2 - The transformation model

Next, I explored the issue of capitalisation in more general terms, i.e., by comparing different patterns of capitalisation both familiar and unfamiliar (e.g., BEERT, beert, beerT). These patterns of familiarity vary by virtue of their presence in the environment. The drastic disruptions provoked by the aLtErNaTiOn CaSe paradigm in visual word recognition were avoided here. A large experiment using the same-different matching task paradigm measured subjects' reaction times to pairs of nonwords exhibiting familiar and unfamiliar patterns of representation. However, only limited case changes were used to differentiate the stimuli. The subject's level of performance varied according to the familiarity of the patterns being matched. The more familiar a physical pattern was the faster subjects matched them. The results show that consistent-case strings were processed faster than the mixed-case strings. Among the consistent strings however, those composed of lower-case letters were processed faster than the upper-case strings. With regards to mixed-case the the initially capitalised pairs were the fastest ones to be processed. As for all the other pairs, with their unfamiliar appearance, no difference in reaction time was observed. The results were discussed in terms of the "orthographic familiarity route" proposed by Besner and Jonhston (1989) model of visual word recognition. The *transformation model* was proposed as explanatory of the internal mechanisms of that route. This hypothesis predicts that there is a hierarchical structure concerning the easiness with which strings are recognised by the cognitive system. The easy with which a string is recognised is dependent upon the level of disturbance that has been caused to its physical appearance and how much this change has caused it to depart from its more familiar shape. The model successfully explains not only the intricate pattern of results of the experiment above, but also can explain other results in the word recognition literature, such as the long time that is taken for subjects to respond to strings that are completely case alternated. A quantified version of the model is provided. To the best of our knowledge this is the first time that such quantified approach is given.

8.1.3 - The vulnerability of initial letter representation

As a contribution to the debate regarding the psychological reality of the representation of letter position in "slot" terms, the neuropsychological case reported by Patterson and Wilson (1990) is revisited and is given a new interpretation. Initial capitalisation at the beginning of a sentence has a different function than that of cueing the reader to the status of proper name. Almost any category of word can be seen initially capitalised at the start of a sentence. Thus, the reader is obliged to cope with the irrelevant variety in the orthography of the initial letter-position by disregarding the initial capitalisation of the word as a signal of its categorical status. In light of the above, I argue that TB's problems originate in a difficulty in sustaining a graphemic representation and relies more on earlier orthographic levels of representation; this produces trouble in integrating the initial letter with the rest of the word, in that TB has effectively traded the orthographic details of the initial letter for graphemic representation, and only imperfect details of the orthography of the initial letter may be retrieved.

8.1.4 - Brand names contribution to familiarity effects

Examining further familiarity effects, a lexical decision and a naming task were used to test the hypothesis put forward by Besner et. al. (1994) concerning the role of figural familiarity in visual word recognition. The innovation here was the use of *brand names* as the stimuli material. There is no report on the literature of this type of stimuli being used before. Previous experiments using acronyms had inherent difficulties, as acronyms (e.g. FBI) are not normally pronounced the same way as common words. So, a tachitoscopic report had to be used, instead of the more straightforward naming task. Brand names, having a more lexicalised pronunciation, are therefore ideal candidates for exploring these questions further. In this study, familiar upper-case brand names and common English words were used as stimuli. Faster reaction times were obtained in the lexical decision task when brand names were presented in their familiar case (capital letters) than when presented in unfamiliar case. No familiarity effect was obtained for common words. In the naming task no familiarity effect was detected either for brand names or common words. The results here confirm that the role of figural familiarity is similar in the categories of brand-names and acronyms. These results are in keeping with Besner's hypothesis which proposes two distinct types of processing in word recognition: the identification process and the figural familiarity process. The identification process uses information that is based only on the knowledge of the abstract identity of the individual letters composing a string. This information is used when the response uniquely specifies a stimulus, as for example, in naming tasks. The figural familiarity process uses information at the word-level, for example, word-shape. It is applied to accomplish tasks such as the lexical decision task that do not demand unique identification. Thus, these results favor the analytic models of word recognition.

8.1.5 - An algorithm for creating nonwords

Nonwords are an important tool in psycholinguistic studies. Also important is the role played by frequency. The traditional method of creating nonwords is to have them derived from real words. Nonword frequency is assumed to be somewhat related to the frequency of the real word from which it has been derived. An alternative method of assigning frequency to nonwords have been created here by means of an algorithm based on the principle that monosyllabic words are composed of sub-syllabic units named onset, nucleus and coda. The frequency of each of these sub-parts was retrieved from a lexical database and this information was used to build the nonwords. All the nonwords used as material in this thesis were built with the help of this algorithm.

8.1.6 - The weird x nonweird strings

The algorithm for creating nonwords related above opened up the opportunity for looking into orthographic familiarity effects with respect to neighborhood density. Two classes of nonwords were created, the *weird* and the *nonweird* nonwords. The nonweird words were defined as those strings which had the combination of its nucleus + coda found in the lexicon. In the weird nonwords this combination was not found in the lexicon. In Experiments 1 and 2, many different pronunciations were found for nonwords which had a more sparse neighborhood (weird group) as compared to the nonweird group. Also, in Experiment 3 the nonweird nonwords were processed faster than the weird. Thus, these results are in line with the majority of findings concerning neighborhood effects reported in the literature.

8.1.7 - The m-score calculation and its applications

Finally, the last study reported in the thesis is that of a computational assessment of the nature of the mappings from letter-to-sound in British English. How reliably can one predict the pronunciation of a string of letters, based solely on statistical information about the grapheme-phoneme relations in English? An exhaustive search of the CELEX database and an analysis of the frequencies of occurrence of grapheme-phoneme correspondences in British English was executed. This information was used to estimate the pronunciation of any string of English graphemes., ie., their *m-scores*. This work is relevant to many different areas of psycholinguistic studies such as spelling, speech and also could be used as a tool to inform studies of the orthographic depth of a language. The algorithm was assessed by examining its behavior for nonwords. This was done by using a corpus of nonword transcriptions which was collected from an experiment with a set of trained phoneticians. The results show that the statistical information about grapheme-phoneme correspondences is not sufficient to predict English pronunciation. More important however, was the failure of the algorithm to supply probabilities for all the transcriptions given by the phoneticians. I suggest that this is due to the fact that other phonologies than the English one were accessed in the attempt of transcribing the pronunciation for unfamiliar strings. A method has also been developed to quantify the orthographical depth of different languages. This quantification is important, for example, in helping researchers in psycholinguistics to know before hand the relationship between two different languages in terms of their orthographic depths before designing and running experiments.

8.1.8 - A re-evaluation of the Search, Logogen and Interactive-activation models in the light of our findings.

In the literature review (Chapter 2), we have talked briefly about the lasting contributions and also the weaknesses of the Search, Logogen and

Interactive-activation models of word recognition. The reason why these particular models were figuring in Chapter 2 is that, although not directly connected to the type of investigation that has been carried out here, in many ways, as explained there, they can be seen as the ancestors of the more contemporary models of word recognition. As for the latter, the hope is that they will, like those models before them, be the framework upon which models of a future generation will be developed. Future models that ultimately will account for human language processing, no matter what idiosyncrasies are part of this processing.

Some of the fundamental concepts that were put forward by those earlier models are still part of the more contemporary ones. For example, all the three earlier models mentioned above, had assumed that feature analysis is the first step towards visual recognition and so do most of the connectionist models around at the moment. Also the concept of "activation" is still with us, even, for example, in the sophistication of the attractor networks (Plaut et. al., 1996). Some of the problems faced by the earlier model are still being faced by their modern counterparts as well. The difference now, lies perhaps, in the recent shift to a more computational approach to modeling. The important consequence of this shift is that it allows researchers to check quantitatively the feasibility of the approach being taken. The issue of nonword processing is a good example for the value of this trend of designing models that can be quantitatively verified. One of the first models to encounter problems in accounting for nonword processing was the Search model. According to its mechanism the rejection of a nonword by the model would only be possible after an exhaustive search through the whole lexicon. The cost of this search is too expensive to match any of the actual empirical data. Also, Seidenberg and McClelland's (1989) model of word recognition was extensively praised for its results regarding the processing of real words in the lexicon, it was however, bitterly criticized for its performance in relation to pseudowords. More recently however, Plaut et. al. (1996) succeeded in building a connectionist network which has its nonword output comparable to the human performance for the range of nonwords chosen (see section 4.3.3.2).

Perhaps, now is the moment to move into the investigation of more fine grained aspects of visual word recognition. This is exactly, what we propose in this thesis, for example, when investigating the influence of initial capitalisation in nonword pronunciation. Unfortunately, it is fair to say that, this more fine grained approach to word recognition was not a concern for the three early models considered above. The issues surrounding this approach are somewhat alien to those models. However, we will attempt here some discussion of the subject in broad terms. Next, we offer some suggestions of how to modify these models so that they could account for the type of behaviour we have found with our experiments. Firstly however, it is important to note that, for the reason just given above, this will be only an exercise of data accommodation, since as it will be seen, no novel explanation was generated out of this effort.

Let us take for example, the Search model. The issue of "word shape" is completely ignored by this model and also by the two other models as well. However, trying to limit the search efforts for a word in the lexicon, Foster has suggested the search to be based on sensory characteristics, such as the visual input, for example, that a word starts with the letter *b* and ends with the letter *e*. We suggest here that perhaps, a further way of constraining the search algorithm would be to incorporate into the model the additional information that if the letter *b* is initially capitalised, the search should be restricted only to other initially capitalised words to be found in the lexicon. The first consequence for the model of adopting this device is to the double the size of its main lexicon, that now, has to store two distinct representations of each word: one initially capitalised and other represented in complete lower case. This is so because, not only proper names are initially capitalised in English, but also any word in the lexicon that is situated at the start of a sentence. Therefore, we suggest that to make the search efficient a further change has to be made to the model and that is to add a top level syntactic unit that will function by telling the model to ignore the vast majority of the initially capitalised words in its search for a proper name, if these are located at the start of a sentence. Ad it can be seen, because the Search model was not designed to accomplish this type of processing, with all the amendments the model becomes cumbersome allowing us to doubt how much of efficiency has really being built in it, after the adaptations.

The Interactive-activation model was built with their authors having in mind to account for the word-superiority effect. In fact, they succeeded in doing so very easily by using the feedback mechanism between the two higher layers composing the model, that is, the letter and the word level. The fact that the word superiority effect could be emulated by the Interactive-activation model lent support to the view that words are recognised letter by letter and not as wholes. However, the contribution of the Interactive-activation model to the issues we have been investigating stops here. In the same fashion as the Search and the Logogen model there is no attempt to account in this model for "word shape" effects. This is of course, partly reflecting the widespread belief in the psycholinguistic community that "word shape effects" do not play any role in visual word recognition. However, as we have shown in many of the experiments reported here, this is not a closed matter and should be further investigated. If next, our results are confirmed, then we suggest that both mechanisms, the one we have suggested in the discussion of Chapter 4 and the one put forward with the transformation model in Chapter 5 should be scrutinized more carefully.

8.2 - Methodological implications and further research

I have argued that in written English, initial capitalisation plays the role of cueing the reader to the category of proper names. If proper names are processed differently from common words, (as has been suggested by a number of researchers, see Valentine et. al, 1996), than further investigation into such visual cues is worth pursuing beyond the present work. The findings inform theoretical modeling of visual word recognition and connectionist implementation and imply the need for some revision of the theories and their implementations. For example, I propose that distributed models should incorporate into their structures, means of restraining their process of generalisation into a restricted lexical space, so that grouped items are all exemplars bearing phonologies that are plausible to that domain. Furthermore they should find a way of incorporating mechanisms that deal with specific aspects of specific categories. In the case of proper names, the processing of visual cues are an integral part of manipulating proper names. Its relevance also extends to the functional models of combined face, name and word recognition (Valentine et. al., 1991).

The investigation reported here relates only to normal processing. However, I suggest that further insight could be gained if future research focused on brain injured subjects. Patients suffering from neglect dyslexia, for example, are potentially interesting targets, since the type of errors they tend to commit are localized on the left side of words. This is precisely the locus of the capitalisation effect. These include different types of errors, such as omissions, substitutions or transpositions of letters and in some cases even whole words are involved.

The complex patterns of results obtained in Experiment 3, described in chapter 5, also offer the possibility for further research. First, further experimentation, probably using the same type of experiment, is desirable so as to verify other conditions such as the manipulation of capitalisation at the centre of the string. Also, different experiments using materials other than nonwords and different task paradigms can also throw more light on the findings obtained here. The implications of these findings hinge on a number of issues involving, for example, the psychological reality of letter position "slots" which is connected to the phenomenon of positional dyslexia described by Katz & Sevush (1989). They are also of interest to recent connectionist models of visual word recognition that take into consideration the fact the exterior letters of words have a privileged psychological representation (Shillcock & Monaghan, submitted).

Another line of research worth pursuing is that of exploring more thoroughly brand names as experimental materials in a wider range of paradigms that could involve for example, research on memory with normals and also impaired subjects. Furthermore, as already mentioned in Chapter 6, brand names offer a very rich environment for those interested in studying the various features of perception, such as colour, size, and font that are an inherent part of their identity and can be investigated using similar methods to those used here.

There is also the algorithm that was created for building nonwords controlled according to the frequency of subsyllabic units. The experimental results discussed here have demonstrated that subjects are sensitive to the frequency of these units. As discussed in Chapter 2, there are some researchers who now believe, that the relations between spellings and sounds in the simple monomorphemic words of English are more predictable when the level of onsets and rimes is taken into account than when only graphemes and phonemes are considered (Treiman et. al., 1995; Bowey, 1996). Therefore, the algorithm above could be very useful in the design of further experiments addressing these concerns.

Finally, the combined wealth of the statistical information on grapheme-phoneme correspondences in British English that has been gathered here together with the program that was developed for predicting pronunciation has resulted in a useful research tool that could be used in different areas of psycholinguistics. An example of that, is the m-score method for calculating the orthographic depth pertaining to different languages that was developed here.

8.3 - Conclusion

Although familiarity effects are a much discussed issue in visual word recognition, the specific processes that underlie the familiarity effects are not yet fully understood. The aim of this thesis has been to demonstrate the existence of an additional number of familiarity effects that until now have been overlooked. A variety of methods were used to achieve this goal. The research reported here has shown the importance of environmental influence in visual word recognition. These results are relevant not only to theoretical approaches to language processing but also to the implementation of connectionist models of visual word recognition.

Appendix I

Lists of materials used in Experiment 1

High-frequency nonwords	*Low-frequency nonwords*
thout	sprouv
hean	screarg
mees	sneertz
toond	shrooln
baif	phaikh
seid	spleij
fier	zieltz
woill	phroitsch
whuim	ghuisch
loay	sphoarrs
nueld	gnuentz
yauth	khaurrh
coew	kroengst
duast	psuavs
sheoch	chreolc
geup	czeurv
ruoght	thwaonst
paoby	phliohn
friont	ptaamf
byaeck	schanaeldst
stawn	rhadz
myeng	dhewm
jits	djiwst
chomy	svowp
knyl	shhulpt

gryk
plaass

schmyhr
kvuochms

Appendix II

Lists of materials used in Experiment 2

Nonweird nonwords	Weird nonwords
bleil	breirn
freing	dreips
heish	keirm
reight	shreig
neich	fleip
seid	treiff
gluid	whuim
fruild	pruigs
guisch	kuigh
sluint	gnuir
bluern	bueck
stued	nueld
puell	huerb
truels	huerth
mieth	luemp
fier	fruech
flauls	rielb
vaugh	giech
plaud	shaup
maut	thauf
creeth	yauth
mees	slaub
choils	splaug
poing	waurst
poid	gnaulp

goils
kaish
saing
thaim
baif
smook
sooch
geuch
pleuch
pleunt
heums
toess
goets
frits
jits
beash
heast
kneath
seath
speard
speash
hean
bours
houst
shuch
knoup
thout
brelp
screlt
screns
tresks
brish
thich
fryss

kaund
woill
shoirp
hoisks
hoirp
smaich
spaing
thaips
waing
spheuk
teucks
meurf
preurb
teup
meuft
wreest
smeent
choec
noerds
sploec
gnoerk
dreagn
keagn
shoib
zoind
thoird
froirn
spaups
strauls
knoong
choong
stuelk
vuest
juelp

scyst
shych
styss
sluill
gluip
brouch
boust
leunt
leuch
feams
bruds
shearn
pueg
fruems
frueb
wruik
kuigh
gnuir
scuiss
moech
hoess
kneull
deugh
teung

Appendix III

Pairs of nonwords used in Experiment 3

Feall - faell
Neust - neuzt
Geock - geuck
Beert - beart
Moick - moyck
Soewn - soeun
Leutch - leutsh
Wraibs - wraibs
Greach - graech
Kniock - knyock
Goakt - goact
Goets - geots
Brysh - brysh
Fruild - fruyld
Luanf - luamf
Thuawn - thuaun
Knoong - knoomg
Traeb - treab
Keirm - keyrm
Noerds - neords
Hiurl - hiarl
Stroiw - stroyw
Hieng - hiemg
Kaink - kaynk
Siemf - sienf

Thoimg - thoing
Shaeng - shaemg
Haerb - haarb
Doumf - dounf
Freing - frieng
Saing - seing
Scyst - scist
Biots - biods
Sluint - sluimt
Poing - poimg
Treah - traeh
Bralp - brelp
Pralp - prulp
Heysh - heish
Shich - shych
Neich - neych
Stiss - styss
Bleil - bleyl
Gaerf - gearf
Griots - griods
Teuck - touck
Sploeb - sbloeb
Theols - teols
Gnaulp - gmaulp
Heeng - heang

Heast - heazt
Preurb - prewrb
Plaup - pleup
Hoorth - hourth
Hiens - hiems
Hiurn - huirn
Screms - screns
Lealb - laelb
Stawn - staun
Knoup - knuop
Fryss - fryzs
Guish - giush
Bours - bours
Shuant - shuamt
Shael - shaal
Shayrn - shairn
Gions - gioms
Sluach - sluash
Svouf - svowf
Phoint - phoimt
Slaels - slaals
Duish - diush
Naess - neass
Plienk - pliemk
Slaub - slawb

Triews - trieus
Liels - lyels
Nialls - nyalls
Smaeg - snaeg
Hoisks - hoysks
Faess - faass
Throip - throyp
Shreus - shrues
Heeld - heald
Bloint - bliont
Thraut - thruat
Striel - stryel
Treich - triech
Bloils - bloyls
Whyong - whyomg
Braunt - braumt
Slymd - slind
Neuld - nueld
Breirn - briern
Pruygs - pruygs
Ptaemf - ptaamf
Smeogs - sneogs
Seond - seomd
Thaalb - thaelb
Twailt - tweilt
Laesh - laash
Horst - hourst
Stoarb - stoab
Gleiw - gliew
Tairb - tayrb
Dreips - dreibs
Draagn - draegn
Naelb - naalb
Trasks - trazks
Shaop - shaup
Fleip - fleyp
Sneegs - sneigs
Dreugs - drewgs
Keads - keeds
Griant - griamt
Knuets - knueds
Thoils - thoyls
Duint - duimt
Miots - miods
Twoink - twoimk
Gnoywn - gnoiwn
Geart - gaert
Giesh - giech
Thaew - thaaw
Proiws - proius
Slaalt - slaelt
Draemp - draamp
Splenk - splemk
Throag -throug
Spheuk - sfheuk
Spaect - sfaect
Thiond - thiomd
Shoum - shuom
Kriont - kryont
Naeng - naang
Traag - traeg
Haong - haomg
Wreaft - wreeft
Screup - scruep
Thaibs - thaips
Pteurd - pdeurd
Daiwn - daiun
Houst - houzt
Heaght - heagt
Piech - piesh
Mirst - myrst
Maewd - maawd
Smeic - smaic
Beath - beeth
Saath - seath
Spescs - spesks
Blascs - blasks
Reyght - reyght
Feath - feeth
Pteeld - ptield
Greit - greyt
Neild - neyld
Heirch - heyrch
Hoump - huomp
Thaim - theim
Droim - droem
Cloups - cluops
Nuerb - noerb
Hiulls - huills
Slasks - slascs
Shraeb - shraab
Thoar - thaor
Heif - haiff
Dryap - driap
Tieng - tiemg
Teeld - teald
Staurt - sdaurt
Faeng - faemg
Duack - dueck
Diech - diekh
Neaps - neabs
Hyorm - hiorm

Faach - faech
Speesh - speash
Thich - thych
Thauzt - thaust
Shais - sheis
Diost - dyost
Truews - trueus
Goiff - goyff
Kaild - kayld
Hiald - hyald
Sphiun - sphuin
Saeft - saevt
Tiogh - tyogh
Mieth - miedh
Hairf - hayrf
Heaus - heaws
Hioft - hiovt
Traokt - traoct
Sheuks - sheucs
Flauf - flaof
Neught - neugt
Hielk - heilk
Bueck - baeck
Tuack - twack
Gnonds - gnomds
Leuch - leush
Brour - brour
Broush - brouch
Kayth - kaith
Speerd - speard
Screlt - skrelt
Whuast - whaust
Knais - knays
Frairf - fruirf

Theegn - theagn
Kneath - kneeth
Knuob - kmuob
Knoit - knoyt
Spoang - spoeng
Haowl - hoawl
Rielb - reelb
Hialb - hyalb
Hiolk - hoilk
Droibs - droips
Teucks - tuecks
Threw - treiw
Smiek - sniek
Chuint - chuimt
Faorm - foarm
Thauf - thouf
Shuch - shach
Gaess - gaass
Shaass - shaess
Chaert - cheart
Froall - fraoll
Fryoch - frioch
Glaurn -gluarn
Ploeds - ploets
Traelt - trealt
Froubs - froups
Waurst - waurzt
Thraec - thraac
Meurf - muerf
Fliot - flyot
Haaps - haeps
Splaug - sflaug
Scries - skries
Treic -triec

Luarf - laurf
Hualf - haulf
Niung - niumg
Shoyrp - shoirp
Thaet - thaat
Shyft - shiff
Thruaw -truaw
Cluirt - cluyrt
Whiest - whiezt
Thuill -thuyll
Grael - greal
Caang - caamg
Smaich - smeich
Chuech - chuesh
Freunt - freumt
Ghaaps - ghaeps
Ptuall - ptaull
Spaall - spaell
Suong - suomg
Deych - deich
Bloip - bloyp
Thruac - truac
Treyw - treiw
Loast - loest
Liung - liumg
Gniac -gnyac
Hioss - hyoss
Treyff - treiff
Luemp - luenp
Shreig - shreyg
Luoms - luons
Giuts - giuds
Hyord -hiord
Ptelb - pdel

Appendix IV

List of materials used in Experiments 4 & 5

Brand names

Name	familiarity	letters	syllables
ASDA	6.2	4	2
ASTRA	5.6	5	2
BOLD	5.7	4	1
CASIO	5.1	5	2
CHANEL	5.6	6	2
COMET	5.7	5	2
DEBENHAMS	6.4	9	3
DRAMBUIE	6.2	8	3
DULUX	5.6	5	2
DURACELL	6.1	8	3
FAIRY	6.1	5	2
FIAT	6.0	4	2
FLORA	6.1	5	2
GUINNESS	6.3	8	2
HEINZ	6.0	5	1
HOLSTEN	6.1	7	2
HONDA	6.0	5	2
HOVIS	5.8	5	2
HYUNDAI	5.5	7	3
LEGO	6.2	4	2
MARMITE	6.2	7	2
MAZDA	5.4	5	2
NESCAFE	6.3	7	3
NISSAN	6.2	6	2
PENTAX	5.5	5	2
PHILIPS	5.6	7	2
PRINGLES	5.4	8	2

Name	familiarity	letters	syllables
QUAKER	5.6	6	2
ROVER	6.2	5	2
SAAB	6.0	4	1
SAFEWAY	6.2	7	2
SENSODYNE	5.8	9	4
SHARP	5.1	5	1
SMIRNOFF	6.4	8	2
SONY	6.1	4	2
SPEEDO	5.2	6	2
SUBARU	5.1	6	3
SUZUKI	6.0	6	3
TESCO	6.3	5	2
THRESHER	5.7	8	2
TOBLERONE	6.0	9	4
TOSHIBA	5.8	7	3
TWININGS	6.0	8	2
VISA	6.2	4	2
VOGUE	5.2	5	2
VOLVO	6.4	5	2
WALKERS	5.4	7	2
YAMAHA	5.8	6	3

Appendix V

Common English Words

Name	letters	syllables	frequency[1]
AMONG	5	2	284
ATTENTION	9	3	300
BANK	4	1	134
BLOOD	5	1	141
CENTURY	7	2	182
CHAPTER	7	2	113
CHILDREN	8	2	655
CITY	4	2	198
COUNCIL	7	2	101
DAUGHTER	8	2	101
DOUBT	5	1	153
EVENING	7	3	183
FAMILY	6	3	328
GARDEN	6	2	113
HEAT	4	1	117
HISTORY	7	2	187
HOSPITAL	8	3	105
HOTEL	5	2	125
INDUSTRY	8	3	146
KITCHEN	7	2	105
LEVEL	5	2	180
LOCAL	5	2	248
MAJOR	5	2	188
MARKET	6	2	133
MATTER	6	1	390
MONEY	5	2	404
MUSIC	5	2	134
PAPER	5	2	174
PARENTS	7	3	246

[1] The frequency (per million) was taken from the CELEX database of English (version 2.5).

Name	letters	syllables	frequency
PARTY	5	2	372
PERSON	6	2	228
PLAY	4	1	275
POLICE	6	3	211
PRESIDENT	9	3	132
PRIVATE	7	3	193
PROBLEMS	8	2	238
READY	5	2	125
REPORT	6	2	136
RESEARCH	8	2	120
RIVER	5	2	108
ROLE	4	2	117
SEVEN	5	2	119
SIDE	4	2	387
SITUATION	9	4	233
STUDENTS	8	2	200
TABLE	5	2	206
VERY	4	2	478

Appendix VI

Tables referring to chapter 6

Table I

Mixed-brands	Familiarity rate	Pure-brands	Familiarity rate	No.of letters	No. of syllables
VISA	6.2	LEGO	6.2	4	2
FIAT	6.0	SAAB	6.0	4	1
ROVER	6.2	TESCO	6.3	5	2
FLORA	6.1	CASIO	5.2	5	2
FAIRY	6.1	HONDA	6.0	5	2
COMET	5.7	DULUX	5.6	5	2
THRESHER	5.7	PRINGLES	5.4	8	2
WALKERS	5.4	PHILIPS	5.6	7	2
QUAKER	5.6	YAMAHA	5.8	6	3
VOGUE	5.2	SPEEDO	5.2	6	2
BOLD	5.7	HOVIS	5.8	5	2
SHARP	5.1	ASTRA	5.6	5	2

(6.1) - Familiarity rate, no. of letters and number of syllables for pure and mixed brand names

Table II

Number of letters composing brand-name strings					
4 letters	5 letters	6 letters	7 letters	8 letters	9 letters
ASDA	ASTRA	NISSAN	HOLSTEN	DRAMBUIE	DEBENHAMS
BOLD	CASIO	SPEEDO	HYUNDAI	DURACELL	SENSODYNE
FIAT	COMET	SUBARU	MARMITE	GUINNESS	TOBLERONE
LEGO	DULUX	SUZUKI	NESCAFE	PRINGLES	
SAAB	FAIRY	QUAKER	PHILIPS	SMIRNOFF	
SONY	FLORA		SAFEWAY	THRESHER	
VISA	HEINZ		TOSHIBA	TWININGS	
	HONDA		WALKERS		
	HOVIS				
	MAZDA				
	PENTAX				
	ROVER				
	SHARP				
	TESCO				
	VOGUE				
	VOLVO				

Table (6.2) - Number of letters composing brand names. There are 23 brand names in the 4-5 letter group and 23 in the other group.

Table III

Number of letters composing common English words					
4 letters	5 letters	6 letters	7 letters	8 letters	9 letters
CITY	MONEY	GARDEN	CHAPTER	DAUGHTER	PRESIDENT
BANK	PAPER	FAMILY	COUNCIL	CHILDREN	SITUATION
HEAT	LEVEL	MARKET	PRIVATE	PROBLEMS	ATTENTION
VERY	MUSIC	POLICE	KITCHEN	STUDENTS	
PLAY	DOUBT	MATTER	EVENING	INDUSTRY	
SIDE	MAJOR	PERSON	HISTORY	HOSPITAL	
	PARTY		CENTURY	RESEARCH	
	RIVER		GENERAL		
	AMONG				
	SEVEN				
	HOTEL				
	LOCAL				
	BLOOD				
	ALONG				
	READY				

Table (6.3) - Number of letters composing common words used in the experiment

Appendix VII

Grapheme-phoneme correspondences in British English.

Graph and ***Phon*** are the grapheme and phoneme respectively.

Prob is the probability of the grapheme being pronounced as the corresponding phoneme.

Prior Prob is the probability of occurrence of the grapheme in the lexicon.

Example word shows an instance of the grapheme-phoneme correspondence.

Pronunciation gives the pronunciation of the example word, using the CELEX "disk" character set.

Freq. is the frequency of ocurrence in the lexicon, of the example word.

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
a	8	0.0102063	0.055504	various	'v8-r7s	2291
a	@	0.157771	0.055504	about	@-'b6t	6256
a	E	0.0140323	0.055504	many	'mE-nI	17343
a	I	0.0114244	0.055504	manager	'm{-nI-_@R	685
a	{	0.496356	0.055504	had	'h{d	22393
a	Q	0.0768164	0.055504	was	'wQz	97174
a	#	0.0885427	0.055504	after	'#f-t@R	13670
a	$	0.065116	0.055504	all	'$1	61767
a	1	0.0797344	0.055504	taken	't1-k@n	6071
b	b	1	0.0171374	back	'b{k	17657
c	k	0.674248	0.0230766	case	'k1s	6866
c	s	0.295289	0.0230766	cent	'sEnt	5019
c	J	0.000458353	0.0230766	cello	'JE-15	33
c	S	0.0300042	0.0230766	depreciate	dI-'pri-SI-1t	1
d	d	0.994743	0.0417323	day	'd1	13729
d	_	0.00525686	0.0417323	education	"E-_U-'k1-SH	4179
e	i	0.0348649	0.0485794	female	'fi-m1l	1044
e	8	0.00720097	0.0485794	parent	'p8-r@nt	1266
e	j	0.0289551	0.0485794	azaleas	@-'z1-lj@z	11
e	@	0.301728	0.0485794	after	'#f-t@R	13670
e	E	0.336907	0.0485794	get	'gEt	5236
e	I	0.259998	0.0485794	because	bI-'kQz	23626
e	#	0.00158907	0.0485794	clerk	'kl#k	404
e	1	0.00543259	0.0485794	elite	1-'lit	398
e	3	0.023324	0.0485794	her	'h3R	69004
f	f	0.677197	0.0247939	from	'frQm	74843
f	v	0.322803	0.0247939	of	'Qv	540085
g	g	0.71464	0.00990682	again	@-'gEn	13783
g	Z	0.0032733	0.00990682	prestige	prE-'stiZ	284
g	_	0.282087	0.00990682	large	'1#_	5842
h	h	0.97836	0.00815717	house	'h6s	8601
h	NP	0.0216404	0.00815717	hour	'6-@R	2867

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
i	2	0.174884	0.0644354	child	'J2ld	7645
i	i	0.00477445	0.0644354	litres	'li-t@z	128
i	j	0.00488411	0.0644354	million	'mI-lj@n	3518
i	@	0.0421366	0.0644354	possible	'pQ-s@-bP	6034
i	I	0.773321	0.0644354	still	'stIl	15641
j	_	1	0.000381684	just	'_Vst	24308
k	k	1	0.00614261	like	'l2k	31903
l	l	1	0.0267161	old	'5ld	12928
m	m	0.999992	0.0255893	more	'm$R	21873
m	F	8.10481e-06	0.0255893	autism	'$-tI-zF	7
n	n	0.847931	0.0711956	not	'nQt	91464
n	N	0.1446	0.0711956	drink	'drINk	1414
n	H	0.00746949	0.0711956	recent	'ri-sHt	1814
o	5	0.143769	0.0615405	most	'm5st	15347
o	@	0.119538	0.0615405	London	'lVn-d@n	5368
o	u	0.162156	0.0615405	into	'In-tu	35436
o	I	0.0137779	0.0615405	women	'wI-mIn	9169
o	Q	0.363086	0.0615405	body	'bQ-dI	5243
o	U	0.00186967	0.0615405	woman	'wU-m@n	6072
o	V	0.0804439	0.0615405	covers	'kV-v@z	197
o	$	0.113899	0.0615405	story	'st$-rI	2992
o	6	0.00146141	0.0615405	hour	'6-@R	2867
p	p	1	0.017981	people	'pi-pP	26215
q	k	1	2.88281e-05	quite	'kw2t	9866
r	r	0.580833	0.0482312	real	'r7l	5354
r	R	0.419167	0.0482312	car	'k#R	4944
s	s	0.561539	0.0607873	last	'l#st	10816
s	z	0.425323	0.0607873	these	'Diz	22697
s	S	0.00554522	0.0607873	sure	'S$R	4678
s	Z	0.0075933	0.0607873	pleasure	'plE-Z@R	1495
t	t	0.940491	0.06867	time	't2m	32093
t	J	0.0152682	0.06867	picture	'pIk-J@R	1905
t	S	0.0442409	0.06867	initiate	I-'nI-SI-1t	21
u	jU	0.137435	0.017351	argument	'#-gjU-m@nt	1584
u	@	0.0962347	0.017351	until	@n-'tIl	5072
u	u	0.0312794	0.017351	revolution	"rE-v@-'lu-SH	1495
u	w	0.000551547	0.017351	language	'l{N-gwI_	2357
u	E	0.0011159	0.017351	burial	'bE-r7l	163
u	I	0.0209622	0.017351	busy	'bI-zI	1012
u	U	0.0561024	0.017351	full	'fUl	4620
u	V	0.437851	0.017351	but	'bVt	96889
u	ju	0.218468	0.017351	united	ju-'n2-tId	3358
v	v	1	0.0112085	view	'vju	3743
w	w	1	0.0143299	week	'wik	4853
z	s	0.0304951	0.000684957	blitz	'blIts	65
z	z	0.967602	0.000684957	size	's2z	2015
z	Z	0.00190324	0.000684957	seizure	'si-Z@R	60
y	j	0.215624	0.0170292	beyond	bI-'jQnd	2588
y	2	0.159196	0.0170292	try	'tr2	1201
y	i	0.000383635	0.0170292	lycee	'li-s1	16
y	@	0.00671753	0.0170292	analysis	@-'n{-l@-sIs	1101
y	I	0.618079	0.0170292	mysterious	mI-'st7-r7s	470
ae	2	0.240833	0.000103832	maestro	'm2s-tr5	14
ae	7	0.0365245	0.000103832	Israelite	'Iz-"r7-l2t	33

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ae	i	0.164788	0.000103832	aesthetic	is-'TE-tIk	244
ae	8	0.0654872	0.000103832	aeroplane	'8-r@-pl1n	145
ae	@	0.192324	0.000103832	gynaecologist	"g2-n@-'kQ-l@-_Ist	22
ae	E	0.00798973	0.000103832	haemorrhage	'hE-m@-rI_	48
ae	I	0.105293	0.000103832	anaesthesia	"{-nIs-'Ti-zj@	13
ae	1	0.18676	0.000103832	Israeli	Iz-'r1-lI	227
ah	@	0.551047	6.93593e-05	Messiah	mI-'s2-@	87
ah	{	0.0053396	6.93593e-05	Fahrenheit	'f{-r@n-h2t	22
ah	#	0.438061	6.93593e-05	Brahmin	'br#-mIn	21
ah	1	0.00555318	6.93593e-05	dahlias	'd1-lj@z	15
ai	2	0.00265759	0.00200115	Kaiser	'k2-s@R	42
ai	8	0.194485	0.00200115	dairy	'd8-rI	190
ai	@	0.0251249	0.00200115	chieftain	'Jif-t@n	21
ai	E	0.368783	0.00200115	again	@-'gEn	13783
ai	{	0.00131769	0.00200115	plaid	'pl{d	49
ai	I	0.0628049	0.00200115	captain	'k{p-tIn	1138
ai	1	0.344827	0.00200115	detail	'di-t1l	880
ai#e	2	0.0499182	0.000108617	aisle	'2l	129
ai#e	8	0.146618	0.000108617	questionnaire	"kwEs-J@-'n8R	26
ai#e	E	0.000545554	0.000108617	bouillabaisse	'bu-j@-bEs	4
ai#e	1	0.802919	0.000108617	appraise	@-'pr1z	2
air	8	1	0.000167399	airline	'8-l2n	184
aigh	1	1	4.17756e-05	straight	'str1t	1318
ao	6	0.207891	1.95249e-05	Maoist	'm6-Ist	35
ao	$	0.792109	1.95249e-05	extraordinary	Iks-'tr$-dH-rI	873
au	5	0.0131004	0.000339242	chauffeur	'S5-f@R	83
au	6	0.00401747	0.000339242	Frau	'fr6	68
au	@	0.0779039	0.000339242	restaurant	'rE-st@-r~N	574
au	Q	0.0133624	0.000339242	sausages	'sQ-sI-_Iz	109
au	#	0.0269869	0.000339242	aunt	'#nt	541
au	$	0.864629	0.000339242	authority	$-'TQ-r@-tI	1757
au#e	5	0.00487149	0.00044094	mauve	'm5v	60
au#e	Q	0.8	0.00044094	because	bI-'kQz	23626
au#e	#	0.00235176	0.00044094	auntie	'#n-tI	70
au#e	$	0.192777	0.00044094	cause	'k$z	1324
augh	#f	0.424297	0.000164377	laughter	'l#f-t@R	870
augh	$	0.575703	0.000164377	daughter	'd$-t@R	1797
eau	5	0.185312	7.78626e-05	bureau	'bj9-r5	289
eau	ju	0.814688	7.78626e-05	beautiful	'bju-t@-fUl	2075
ee	7	0.0322277	0.00370124	steering	'st7-rIN	131
ee	i	0.872782	0.00370124	need	'nid	3002
ee	I	0.0920123	0.00370124	committee	k@-'mI-tI	2074
ee	1	0.00297782	0.00370124	fiancee	fI-'qN-s1	27
ee#e	7	0.00808733	0.000701563	eerie	'7-rI	80
ee#e	i	0.991913	0.000701563	cheese	'Jiz	497
ei	2	0.0736682	0.00167489	stein	'st2n	788
ei	7	0.00265343	0.00167489	simultaneity	"sI-m@l-t@-'n7-tI	9
ei	i	0.215255	0.00167489	receiver	rI-'si-v@R	251
ei	8	0.464042	0.00167489	heiress	'8-rIs	23
ei	@	0.0239517	0.00167489	foreigners	'fQ-r@-n@z	274
ei	E	0.00137979	0.00167489	leisure	'lE-Z@R	464
ei	I	0.207118	0.00167489	ageing	'1-_IN	136
ei	1	0.0119316	0.00167489	reign	'r1n	105
ei#e	i	0.923117	2.87096e-05	receive	rI-'siv	209

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ei#e	1	0.0768834	2.87096e-05	beige	'b1Z	42
eir	8	1	0.000774715	their	'D8	51922
eigh	2	0.114425	0.00013063	height	'h2t	616
eigh	1	0.885575	0.00013063	eight	'1t	1839
eou	6	0.012464	3.09021e-05	lineout	'l2n-6t	23
eou	7	0.129914	3.09021e-05	hideous	'hI-d7s	185
eou	@	0.857622	3.09021e-05	courageous	k@-'r1-_@s	100
eu	3	0.10397	4.40273e-05	amateurish	"{-m@-'t3-rIS	17
eu	4	0.0726783	4.40273e-05	Freudian	'fr4-d7n	53
eu	7	0.301144	4.40273e-05	museum	mju-'z7m	548
eu	@	0.522207	4.40273e-05	amateur	'{-m@-t@R	204
eu#e	3	0.0316302	6.08857e-06	masseuse	m{-'s3z	1
eu#e	ju	0.96837	6.08857e-06	queue	'kju	178
ey	2	0.0676855	0.00180302	geyser	'g2-z@R	20
ey	i	0.0138937	0.00180302	keyhole	'ki-h5l	14
ey	I	0.139142	0.00180302	money	'mV-nI	7226
ey	1	0.779279	0.00180302	surveys	's3-v1z	155
ey#e	2	0.981435	3.51093e-05	eye	'2	2284
ey#e	I	0.0185654	3.51093e-05	honeybee	'hV-nI-bi	4
ieu	j3	0.450262	8.48845e-06	Monsieur	m@-'sj3R	207
ieu	Ef	0.47993	8.48845e-06	lieutenant	lEf-'tE-n@nt	247
ieu	ju	0.069808	8.48845e-06	lieu	'lju	23
iew	ju	1	0.000110142	review	rI-'vju	290
igh	2	1	0.00125965	light	'l2t	4204
oo#e	u	0.81007	4.64865e-05	choose	'Juz	288
oo#e	U	0.18993	4.64865e-05	goodie	'gU-dI	5
ou#e	3	0.000924464	0.000657003	scourge	'sk3_	39
ou#e	6	0.407373	0.000657003	house	'h6s	8601
ou#e	9	0.153844	0.000657003	bourse	'b9s	22
ou#e	u	0.118129	0.000657003	route	'rut	710
ou#e	$	0.319729	0.000657003	course	'k$s	11914
ou	3	0.0045251	0.00990637	journal	'_3-nP	318
ou	5	0.00658577	0.00990637	soul	's5l	731
ou	6	0.362866	0.00990637	around	@-'r6nd	5406
ou	9	0.00888869	0.00990637	tenuous	'tE-nj9s	51
ou	@	0.102437	0.00990637	conscious	'kQn-S@s	807
ou	u	0.221613	0.00990637	group	'grup	5462
ou	U	0.173063	0.00990637	should	'SUd	3552
ou	V	0.0465355	0.00990637	country	'kVn-trI	6036
ou	$	0.0734857	0.00990637	four	'f$R	5843
oo	5	0.00364071	0.00277913	brooch	'br5J	33
oo	9	0.000122601	0.00277913	boorish	'b9-rIS	5
oo	u	0.41745	0.00277913	food	'fud	4552
oo	U	0.465989	0.00277913	good	'gUd	16874
oo	V	0.0253837	0.00277913	blood	'blVd	2536
oo	$	0.0874142	0.00277913	door	'd$R	5891
ough	Vf	0.15512	0.00121878	enough	I-'nVf	5684
ough	5	0.205878	0.00121878	though	'D5	10329
ough	6	0.0296091	0.00121878	ploughing	'pl6-IN	93
ough	@	0.0143427	0.00121878	borough	'bV-r@	96
ough	u	0.231306	0.00121878	through	'Tru	14964
ough	Qf	0.00635696	0.00121878	cough	'kQf	195
ough	$	0.357388	0.00121878	thought	'T$t	2110
ow	5	0.43198	0.00405014	know	'n5	6088

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ow	6	0.442247	0.00405014	down	'd6n	18344
ow	@	0.101683	0.00405014	arrowroot	'{-r@-rut	3
ow	Q	0.0240893	0.00405014	knowledge	'nQ-lI_	2421
ow#e	5	0.62855	1.98212e-05	owe	'5	56
ow#e	6	0.37145	1.98212e-05	browse	'br6z	3
u#e	3	0.0572755	0.00213176	nurse	'n3s	556
u#e	9	0.000903399	0.00213176	lure	'l9R	63
u#e	@	0.24348	0.00213176	future	'fju-J@R	2437
u#e	u	0.0820495	0.00213176	absolute	'{b-s@-lut	755
u#e	I	0.0144057	0.00213176	minute	'mI-nIt	1321
u#e	$	0.0554478	0.00213176	assure	@-'S$R	61
u#e	V	0.126198	0.00213176	bulge	'bVl_	58
u#e	ju	0.42024	0.00213176	abuse	@-'bjus	253
bb	b	1	0.000136304	rubble	'rV-bP	107
bt	t	1	8.36845e-05	subtle	'sV-tP	393
cch	k	1	2.66653e-07	saccharine	's{-k@-r2n	6
cc	k	1	0.000527261	accurate	'{-kjU-r@t	377
cht	t	1	4.38496e-06	yacht	'jQt	83
cqu	kw	0.979292	2.71838e-05	acquaintance	@-'kw1n-t@ns	208
cqu	k@	0.0207084	2.71838e-05	lacquer	'l{-k@R	17
dd	d	1	0.000417326	middle	'mI-dP	1642
gg	g	0.698684	0.000300592	aggressive	@-'grE-sIv	385
gg	_	0.301316	0.000300592	suggest	s@-'_Es-t	338
gh	g	1	2.3821e-05	ghost	'g5st	351
mm	m	1	0.000678839	committee	k@-'mI-tI	2074
nn	n	1	0.000634367	innocence	'I-n@-s@ns	316
ph	f	0.987801	0.000439592	telephone	'tE-lI-f5n	1738
ph	v	0.00542563	0.000439592	nephew	'nE-vju	136
ph	p	0.00677361	0.000439592	shepherds	'SE-p@dz	68
pp	p	1	0.00122494	apple	'{-pP	315
pph	f	1	5.92562e-07	sapphire	's{-f2-@R	25
qu	kw	0.910513	0.00109442	adequate	'{-dI-kw@t	540
qu	k	0.0894866	0.00109442	technique	tEk-'nik	567
rr	r	0.996158	0.000936929	terrible	'tE-r@-bP	1391
rr	R	0.00384214	0.000936929	bizarre	bI-'z#R	182
tsch	J	1	7.77737e-06	putsch	'pUJ	8
sch	s	0.958675	0.000198597	schism	'sI-z@m	15
sch	S	0.0413248	0.000198597	schedule	'SE-djul	231
tch	J	1	0.000332546	dispatch	dI-'sp{J	270
tt	t	1	0.00175875	attention	@-'tEn-SH	2369
wh	h	0.203688	0.00438359	who	'hu	42881
wh	w	0.796312	0.00438359	which	'wIJ	61399
zz	z	0.957131	3.49019e-05	puzzled	'pV-zPd	237
zz	ts	0.0428693	3.49019e-05	mezzanine	'mE-ts@-nin	2
ci	S	1	0.00068509	social	's5-SP	7192
ck	k	1	0.00160214	back	'b{k	3090
th	t	0.00703618	0.0344971	thirdly	't3d-lI	144
th	D	0.883223	0.0344971	other	'V-D@R	29391
th	T	0.109741	0.0344971	nothing	'nV-TIN	9026
ch	k	0.104482	0.00428378	character	'k{-r@k-t@R	1569
ch	J	0.869437	0.00428378	each	'iJ	11320
ch	S	0.0203963	0.00428378	machine	m@-'Sin	1469
ch	_	0.00568524	0.00428378	sandwich	's{n-wI_	184
cz	z	1	3.70351e-07	czar	'z#R	20

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ct	t	1	2.54802e-06	indictment	In-'d2t-m@nt	87
al	P	0.876058	0.00218024	capital	'k{-pI-tP	1687
al	$	0.123942	0.00218024	talking	't$-kIN	3267
dg	_	1	0.000262801	bridge	'brI_	1034
ce	S	1	8.13291e-06	curvaceous	k3-'v1-S@s	1
di	_	1	2.24729e-05	soldiers	's51-_@z	1028
dj	_	1	1.91842e-05	adjustment	@-'_Vst-m@nt	150
ed	d	0.622442	0.00511695	turned	't3nd	1163
ed	t	0.377558	0.00511695	looked	'lUkt	1824
el	P	1	0.000302443	level	'lE-vP	3224
en	H	1	0.00160814	even	'i-vH	23445
ff	f	1	0.00122368	effect	I-'fEkt	2781
gue	g	1	0.000133489	league	'lig	551
ngue	N	1	1.19105e-05	tongue	'tVN	602
il	P	1	0.000101387	civil	'sI-vP	1209
in	H	1	0.000289318	basin	'b1-sH	270
le	P	1	0.00258421	simple	'sIm-pP	2693
ll	l	1	0.00593793	college	'kQ-lI_	1380
mn	m	0.9969	5.25602e-05	damned	'd{md	197
mn	n	0.00310034	5.25602e-05	mnemonics	ni-'mQ-nIks	6
ol	P	1	4.14645e-05	pistol	'pI-stP	254
on	H	1	0.00102368	person	'p3-sH	4090
pb	b	1	6.20708e-06	cupboard	'kV-b@d	270
pn	n	1	1.30364e-06	pneumonia	nju-'m5-nj@	66
ft	f	1	0.000129193	often	'Q-fH	8409
gi	_	1	0.000554253	regionally	'ri-_H-@-lI	6
gm	m	1	3.1702e-06	paradigm	'p{-r@-d2m	54
gn	n	1	0.000214241	campaign	k{m-'p1n	1438
kn	n	1	0.000760938	know	'n5	6088
kh	k	1	5.08122e-06	khan	'k#n	116
ld	d	1	0.00167562	should	'SUd	3552
lm	m	1	3.38945e-05	calm	'k#m	418
lv	v	1	5.30343e-06	calves	'k#vz	183
mb	m	1	9.38469e-05	climbed	'kl2md	121
ng	N	1	0.00922306	thing	'TIN	9722
li	P	1	5.18936e-05	settling	'sE-tP-IN	178
ly	P	1	2.58801e-05	facially	'f1-SP	5
nd	n	1	5.27232e-05	grandparents	'gr{n-"p8-r@nts	136
ps	s	1	4.92863e-05	psychological	"s2-k@-'lQ-_I-kP	664
pt	t	1	2.97762e-06	receipt	rI-'sit	122
rh	r	1	2.0636e-05	rhythm	'rI-D@m	332
sc	k	0.00853741	0.000131875	viscount	'v2-k6nt	61
sc	s	0.980678	0.000131875	scientific	"s2-@n-'tI-fIk	1119
sc	S	0.0107841	0.000131875	fascism	'f{-SI-z@m	109
sh	S	1	0.00364994	fish	'fIS	1438
si	S	0.454347	0.000462405	version	'v3-SH	673
si	Z	0.545653	0.000462405	vision	'vI-ZH	816
sl	l	1	3.21761e-05	island	'2-l@nd	1209
ss	s	0.888692	0.00257052	possible	'pQ-s@-bP	6034
ss	z	0.0439318	0.00257052	possession	p@-'zE-SH	508
ss	S	0.0673759	0.00257052	issues	'I-Suz	1019
st	s	1	0.00014537	castle	'k#-sP	424
sw	s	1	8.69288e-05	answer	'#n-s@R	1833
tg	g	1	6.62188e-06	mortgage	'm$-gI_	310

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ti	J	0.0535881	0.00308759	question	'kwEs-J@n	5107
ti	S	0.944128	0.00308759	emotions	I-'m5-SHz	540
ti	Z	0.00228382	0.00308759	equation	I-'kw1-ZH	104
tw	t	1	0.000367773	two	'tu	24552
wr	r	1	0.000283748	wrong	'rQN	3078
xh	gz	1	2.26655e-05	exhaustion	Ig-'z$s-J@n	214
x	ks	0.822824	0.001923	experience	Ik-'sp7-r7ns	3389
x	gz	0.164565	0.001923	example	Ig-'z#m-pP	4326
x	kS	0.0126108	0.001923	sexual	'sEk-S91	1443
a#e	8	0.0215976	0.0097077	square	'skw8R	726
a#e	@	0.173475	0.0097077	ultimate	'Vl-tI-m@t	501
a#e	I	0.0395404	0.0097077	heritage	'hE-rI-tI_	188
a#e	{	0.225532	0.0097077	collapse	k@-'l{ps	277
a#e	Q	0.000112925	0.0097077	blancmange	bl@-'mQn_	6
a#e	#	0.16991	0.0097077	morale	mQ-'r#l	181
a#e	$	0.00737368	0.0097077	false	'f$ls	785
a#e	1	0.362297	0.0097077	face	'f1s	7200
a#e	U	0.000119029	0.0097077	immaculate	I-'m{-kjU-lUt	78
a#e	q	4.12023e-05	0.0097077	melange	m1-'lqnZ	10
eaux	5	1	6.2219e-07	plateaux	'pl{-t5	9
oux	u	1	2.37025e-06	Sioux	'su	76
e#e	i	0.738153	0.0335221	complete	k@m-'plit	1553
e#e	8	0.0418564	0.0335221	where	'w8R	11857
e#e	@	0.0500875	0.0335221	dependence	dI-'pEn-d@ns	229
e#e	E	0.0506872	0.0335221	fence	'fEns	399
e#e	I	0.0753472	0.0335221	college	'kQ-lI_	1380
e#e	1	0.00616124	0.0335221	creche	'kr1S	39
e#e	3	0.0280137	0.0335221	reserve	rI-'z3v	313
e#e	7	0.00969394	0.0335221	severe	sI-'v7R	625
i#e	2	0.468732	0.00755894	alive	@-'l2v	1135
i#e	i	0.0235157	0.00755894	police	p@-'lis	3694
i#e	@	0.0459692	0.00755894	Cheshire	'JE-S@R	63
i#e	I	0.461783	0.00755894	bridge	'brI_	1034
o#e	5	0.222629	0.00655299	close	'kl5s	1526
o#e	@	0.146877	0.00655299	purpose	'p3-p@s	1651
o#e	u	0.0282423	0.00655299	move	'muv	580
o#e	V	0.225001	0.00655299	above	@-'bVv	3056
o#e	$	0.199998	0.00655299	store	'st$R	871
o#e	wV	0.177253	0.00655299	one	'wVn	34640
y#e	2	0.356828	0.000242121	type	't2p	1642
y#e	I	0.643172	0.000242121	tricycle	'tr2-sI-kP	11
oy	4	1	0.000344871	joy	'_4	723
oy#e	4	1	1.19549e-05	Joyce	'_4s	210
ua	9	1	0.000383965	actual	'{k-J91	1081
ue	u	0.409877	0.00043617	issue	'I-Su	1564
ue	ju	0.590123	0.00043617	values	'v{l-juz	1037
ui	ju	0.0720752	0.000593999	pursuit	p@-'sjut	371
ui	u	0.194828	0.000593999	juicy	'_u-sI	66
ui	I	0.733097	0.000593999	build	'bIld	311
ui#e	u	1	2.57024e-05	juice	'_us	365
ul	P	1	9.25878e-06	faculty	'f{-kP-tI	422
uo	4	1	2.57764e-06	buoyant	'b4-@nt	62
uy	2	1	5.93895e-05	buy	'b2	565
ya	7	1	3.21465e-06	Aryan	'8-r7n	23

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ar	@	0.135247	0.00479821	afterwards	'#f-t@-w@dz	886
ar	#	0.777463	0.00479821	cards	'k#dz	476
ar	$	0.0828785	0.00479821	awards	@-'w$dz	139
ar	8	0.00441191	0.00479821	stares	'st8z	115
ar#e	#	1	0.000138571	charge	'J#_	1180
are	8	1	0.000110157	awareness	@-'w8-nIs	433
aw	$	1	0.000386098	raw	'r$	669
aw#e	$	1	4.222e-06	awe	'$	148
ay#e	2	0.0831799	4.82641e-05	aye	'2	102
ay#e	1	0.91682	4.82641e-05	aye	'1	122
ay	2	0.00269808	0.00271235	Mayan	'm2-@n	15
ay	8	0.00465337	0.00271235	prayer	'pr8R	360
ay	I	0.0829961	0.00271235	yesterday	'jE-st@-dI	969
ay	1	0.909652	0.00271235	day	'd1	13729
ea	i	0.361016	0.00689072	beach	'biJ	1060
ea	7	0.222318	0.00689072	real	'r7l	5354
ea	8	0.018134	0.00689072	bear	'b8R	227
ea	@	0.0468324	0.00689072	sergeant	's#-_@nt	381
ea	E	0.165423	0.00689072	ahead	@-'hEd	1721
ea	{	0.00203806	0.00689072	whereas	w8-'r{z	867
ea	I	0.12295	0.00689072	forehead	'fQ-rId	438
ea	#	0.00916053	0.00689072	thereafter	"D8-'r#f-t@R	197
ea	1	0.0521275	0.00689072	greatest	'gr1-tIst	1198
ea#e	i	0.502649	0.000606768	leave	'liv	1045
ea#e	7	0.116751	0.000606768	laureate	'l$-r7t	11
ea#e	@	0.240582	0.000606768	vengeance	'vEn-_@ns	128
ea#e	E	0.0789814	0.000606768	cleanse	'klEnz	5
ea#e	I	0.0610366	0.000606768	mileage	'm2-lI_	64
et	1	1	1.32438e-05	beret	'bE-r1	32
eo	5	0.0258747	0.000565659	yeoman	'j5-m@n	34
eo	i	0.701681	0.000565659	people	'pi-pP	26215
eo	7	0.111067	0.000565659	theories	'T7-rIz	557
eo	@	0.0348837	0.000565659	dungeon	'dVn-_@n	24
eo	E	0.018149	0.000565659	leopard	'lE-p@d	118
eo	I	0.0948041	0.000565659	pigeon	'pI-_In	38
eo	Q	0.0135397	0.000565659	thereof	"D8-'rQv	33
ew	5	0.00822859	0.000948765	sewing	's5-IN	73
ew	j	0.00772894	0.000948765	stewardess	"stj9-'dEs	64
ew	u	0.15397	0.000948765	blew	'blu	96
ew	ju	0.823952	0.000948765	few	'fju	2933
ew	jU	0.0061207	0.000948765	steward	'stjU-@d	145
ew#e	ju	1	2.08878e-06	ewe	'ju	22
ia#e	7	0.395906	7.67071e-05	appropriate	@-'pr5-pr7t	1082
ia#e	I	0.604094	7.67071e-05	carriage	'k{-rI_	231
ia	7	0.305669	0.000782359	historian	hI-'st$-r7n	171
ia	@	0.694331	0.000782359	Belgian	'bEl-_@n	126
ie	2	0.15562	0.00220183	lies	'l2z	877
ie	7	0.0909568	0.00220183	Soviet	's5-v7t	2543
ie	i	0.167838	0.00220183	field	'fild	2279
ie	@	0.0683639	0.00220183	gaiety	'g1-@-tI	61
ie	E	0.0578547	0.00220183	friend	'frEnd	3087
ie	I	0.459366	0.00220183	babies	'b1-bIz	1339
ie#e	7	0.314732	0.000367136	inexperience	"In-Ik-'sp7-r7ns	23
ie#e	i	0.387524	0.000367136	believe	bI-'liv	1289

Graph	Phon	Prob	Prior Prob	Example Word	Pronunciation	Freq.
ie#e	@	0.0664165	0.000367136	conscience	'kQn-S@ns	425
ie#e	I	0.231328	0.000367136	sieve	'sIv	38
oa#e	5	0.532508	9.56987e-06	loathe	'15D	9
oa#e	$	0.467492	9.56987e-06	coarse	'k$s	196
oa	5	0.759182	0.000478775	boat	'b5t	1000
oa	$	0.240818	0.000478775	broad	'br$d	751
oe	5	0.270681	0.000312665	goes	'g5z	2727
oe	i	0.0096655	0.000312665	amoeba	@-'mi-b@	22
oe	9	4.73799e-05	0.000312665	wrongdoers	"rQN-'d9z	1
oe	u	0.119871	0.000312665	shoes	'Suz	1168
oe	V	0.597366	0.000312665	does	'dVz	8758
oe	$	0.00236899	0.000312665	Boer	'b$R	30
oea	7	1	1.42215e-06	diarrhoea	"d2-@-'r7	92
oh	5	0.64892	0.000190597	ohms	'5mz	12
oh	Q	0.35108	0.000190597	John	'_Qn	3566
oi	4	0.980339	0.000470924	avoid	@-'v4d	337
oi	w#	0.0196609	0.000470924	reservoir	'rE-z@-vw#R	81
oi#e	4	0.967284	0.000116824	choice	'J4s	1828
oi#e	w#	0.0327162	0.000116824	armoire	'#m-w#R	5
iu	7	1	1.98508e-05	equilibrium	"i-kwI-'lI-br7m	122
iou	7	1	0.000274104	glorious	'gl$-r7s	223
io	7	0.57145	0.000193442	idiot	'I-d7t	173
io	@	0.42855	0.000193442	passionate	'p{-S@-n@t	279
ir	3	1	0.000873347	bird	'b3d	752
er	3	0.216614	0.017134	alternative	$1-'t3-n@-tIv	542
er	7	0.0582706	0.017134	adhered	@d-'h7d	10
er	8	0.0843807	0.017134	concerto	k@n-'J8-t5	79
er	@	0.625347	0.017134	clusters	'klV-st@z	109
er	#	0.00227044	0.017134	clerk	'kl#k	404
er	1	0.0131177	0.017134	dossier	'dQ-sI-1	76
er#e	3	0.0999564	0.00210748	adverse	'{d-v3s	132
er#e	8	0.900044	0.00210748	there	'D8R	60143
re	@R	1	0.00592175	centre	'sEn-t@R	3123
ern	H	1	0.00013666	govern	'gV-vH	30
ear	3	0.287408	0.00151543	earthquake	'3T-kw1k	97
ear	7	0.673901	0.00151543	fearsome	'f7-s@m	37
ear	#	0.0386913	0.00151543	heart	'h#t	2597
eer	7	1	0.000121031	cheers	'J7z	55
or	3	0.0800582	0.0111778	work	'w3k	747
or	@	0.127223	0.0111778	effort	'E-f@t	1567
or	R	0.256235	0.0111778	mayor	'm8R	268
or	$	0.536484	0.0111778	former	'f$-m@R	1177
or#e	$	1	0.000114335	force	'f$s	2436
uar	#	1	3.26353e-05	guard	'g#d	658
ur	3	0.68595	0.000926248	absurd	@b-'s3d	450
ur	9	0.0242303	0.000926248	lured	'19d	9
ur	@	0.28982	0.000926248	Saturday	's{-t@-dI	1037
ur#e	3	1	0.000128171	curve	'k3v	276
our	3	0.0684091	0.000608724	journey	'_3-nI	923
our	9	0.00710618	0.000608724	tourney	't9-nI	1
our	@	0.251564	0.000608724	colours	'kV-1@z	511
our	$	0.672921	0.000608724	court	'k$t	2228
lf	f	1	0.000106513	half	'h#f	3065
eye	2	1	0.000155651	eye	'2	2284

References

Adams, M. J. (1979). Models of word recognition. *Cognitive Psychology*, 11, 133-176.

Adams, M. (1990). Beginning to read. Cambridge, MA:MITPress

Albrecht, J. E., & O'Brien, E. J. (1991). Effects if centrality on retrieval of text-based concepts. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 932-939.

Allport, D.A. (1979). Word recognition in reading: A tutorial review. In P. A. Kolers, H. Bouma, and M. Wrolstad (Eds.), *Processing of Visible Language*, vol.1. New York: Plenum.

Andrews, S. (1989). Frequency and neighboorhood effects on lexical access: Activation or search. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 802-814.

Andrews, S. (1992). Frequency and neigborhood effects on lexical access: lexical similarity or orthographic redundancy ? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 802-814.

Baayen, R. H., Pipenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pensylvania, Philadelphia, PA.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access ? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 340-357.

Baron, J. (1977). Mechanisms for pronouncing printed words: use and acquisition. In D. Laberge and S. J. Samuels (Eds.), *Basic processes in reading: perception and comprehension*. London: LEA.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 386-393.

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, 4, 207-228.

Barron, R. W., & Pittenger, J.B. (1974). The effect of orthographic structure and lexical meaning on "same-different" judgements. *Quartely Journal of Experimental Psychology*, 26, 566-581.

Beauvois, M.-F., & Derouesne, J. (1979). Phonological alexia: *Three dissociations. Journal of Neurology, Neurosurgery and Psychiatry*, 42, 1115-1124.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556-566.

Bennet, D. M. (1967). New Methods and Materials in Spelling. Victoria, Australian Council for Educational Research.

Berndt, R.S., Reggia, J.A., & Mitchum, C.C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior, Research Methods, Instruments &Computers*, 19, 1-9.

Besner, D. (1980). Visual word recognition: codes and procedures for accessing the internal lexicon. Unpublished Ph.D. thesis, Reading University, England.

Besner, D. (1983). Basic decoding components in reading: Two dissociable feature extraction processes. *Canadian Journal of Psychology-Revue Canadienne De Psychologie*, 37, 429-438.

Besner, D., & Johnston, J. C. (1989). Reading and the mental lexicon: On the uptake of visual information. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 291-316). Cambridge: MA: MIT Press.

Besner, D., & McCann, R.S. (1987). Word frequency and pattern distortion in visual word identification and production: An examination of four classes of models. In M.Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*. Hillsdale, N.J.:Erlbaum.

Besner, D., Davelaar, E., Alcott, D., & Parry, P. (1984). Wholistic reading of alphabetic print: Evidence from the FDM and the FBI. In L. Henderson (Ed.), *Orthographies and reading: Perspectives from cognitive psychology, neuropychology and linguistics*. Hillsdale,NJ: Lawrence Erlbaum Associates.

Besner, D., Twilley, L., McCannan, R.S., & Seergobin, K. (1990). On the association between connectionism and data - are a few words necessary ? *Psychological Review*, 97, 432-446.

Bolla, K.I., Lindgren, K.N., Bonaccorsy, C., & Bleecker, M.L. (1991). Memory complaints in older adults, fact or fiction ? *Archives of Neurology*, 48, 61-64.

Bradley, D.C., & Garret, M.F. (1983). Hemispheric differences in the recognition of closed and open class words. *Neuropsychologia*, 21, 155-159.

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read - A causal connection. *Nature*, 301, 419-421.

Brennen, T. (1993). The difficulty with recalling people's names: The plausible phonology hypothesis. *Memory*, 409-431.

Brennen,T., Baguley, T., Bright, J., & Bruce, V. (1990). Resolving semantically induced tip-of-the-tongue states for proper nouns. *Memory and Cognition*, 18, 339-347.

Broca, P. (1861). Remarques sur le siège de la falcuté du langage articulé suivie d'une observation d'aphémie. *Bull. Soc. Anat. Paris*, 6, 330. [Translated in R. Herrnstein & E.G. Boring, (1965). *A source book in the history of psychology*. Cambridge, MA: Havard Universtity Press.

Brooks, L. (1977). Visual pattern in fluent word identification. In Reber, A.S., & Scarborough, D.L., *The Proceedings of the Cuny Confererences*. London, LEA.

Brow, R.W., & McNeill, D. (1966). The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behaviour*, 5, 325-337.

Brown, G. D. A., & Ellis, N. C. (1994). Issues in spelling research: an overview. In Brown, G.D.A & Ellis, N. C. (Eds.), *Handbook of Spelling*, Chichester, Wiley, & Sons.

Bruder, G. A. (1978). Role of visual familiarity in the word-superiority effects obtained with simultaneous-matching task. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 88-100.

Burke, D.M., MacKay, D.G., Worthley, J., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults ? *Journal of Memory and Language*, 30, 542-579.

Burke, D.M., Worthley, J., & Martin, J. (1988). I'll never forget what's her name: Aging and the tip of the tongue experiences in everyday life. In M.M. Gruneberg, P.E. Morris and R.N. Skyes (Eds.), *Practical aspects of memory: current research and issues*, vol.2. Chichester:Wiley & Sons.

Burton, A. M., & Bruce, V. (1992). I recognise your face but I can't remember your name: A simple explanation ? *British Journal of Psychology*, 83, 45-60.

Burton, A.M., & Bruce, V. (1993). Naming faces and naming names: Exploring an interactive activation model. *Memory*, 1, 457-480.

Burton, A.M., Bruce, V., & Johnston, R.A. (1990). Understanding face recognition with an interactive model. *British Journal of Psychology*, 81, 361-380.

Campbell, R., & Besner, D. (1981). This and Thap - constraints on the pronunciation of new written words. *Quartely Journal of Experimental Psychology*, 33A, 375-396.

Caramazza, A., & Berndt, R.S. (1978). Semantic and syntactic process in aphasia: A review of the literature. *Psychological Bulletin*, 85, 898-918.

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind*, 11, 220-242.

Chambers., S. M., & Forsters., K .I. (1975). Evidence for lexical access in a simultaneous matching task. *Memory & Cognition*, 3, 549-559.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cohen, G. (1990). Why is it difficult to put names to faces ? *British Journal of Psychology*, 81, 287-297.

Cohen, G., & Burke, D.M. (1993). Memory for proper names: A review. *Memory*, 4, 249-263.

Cohen, G., & Faulkner, D. (1984). Memory in old age: "Good in parts". *New Scientist*, 11, 49-51.

Cohen, J.D., Mac Whinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers*, 25, 257-271.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed). *Strategies of information processing* (pp. 151-216). London: Academic Press.

Coltheart, M. (1981). Disorders of reading and their implications for models of normal reading. *Visible Language*, 15, 245-286.

Coltheart, M., & Freeman, N. (1974). Case alternation impairs word identification. *Bulletin of Psychonomic Society* 3,102-104.

Coltheart, M., Davelaar, E., Jonasson, J.T. & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* (pp 535-555). Hilsdale, NJ: Erlbaum.

Coltheart, M., Patterson, K., & Marshall J. C. (Eds.) (1980). *Deep Dyslexia*. London: Routledge & Kegan Paul.

Czerniewska, P. (1992). Learning about writing. Blackwell, Oxford.

DE Groot, a. M. B. (1984). Primed lexical decision: combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *Quartely Journal of Experimental Psychology*, 36A, 253-280.

Driver, J., & Baylis, G. C. (1995). Tilted letters and tilted words: A possible role for principal axes in visual word recognition. *Memory and Cognition*, 23, 560-568.

Dunbar, K., & MacLwod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 622-302.

Dyer, F. N. (1971). Color naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 10, 297-302.

Egeth, H., & Blecker, D. (1971). Differential effects of familiarity on judgements of sameness and difference. *Perception and Psychophysics*, 9, 321-326.

Eichelman, W.A. (1970). Familiarity effects in the simultaneous matching task. *Journal of Experimental Psychology*, 86, 275-282.

Ellis, H. D. & Young, A. W. (1992). Age of acquisition and recognition of nouns presented in the left and right visual fields: A failed hypothesis. *Neuropsychologia*, 15, 825-828.

Ellis, N. C. (1994). Longitudinal studies of spelling development. In Brown, G. D. A., & Ellis, N. C. (Eds.), *Handbook of Spelling*, Chichester, Wiley, & Sons.

Erdmann, B., & Dodge, R. (1898). *Psychologische Untersuchungen uber das Lesen* (Psychological research in reading). Halle, GDR:Niemeyer.

Evett, L., & Humphreys, G.W. (1981). The use of abstract graphemic information in lexical access. *Quartely Journal of Experimental Psychology* 33A, 325-350.

Fery, P., Vincent, E., & Bredart, S. (1995). Personal anomia: A single case study. *Memory and Cognition*, 5, 335-339.

Feustel, T.C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General* , 112, 309-346.

Fisher, D. F. (1975). Reading and visual search. *Memory and Cognition*, 3, 188-196.

Fit, S. (1995). The pronunciation of unfamiliar native and non-native town names. *Proceedings of the Eurospeech Conference*, vol.3, pp. 2227-30.

Fit, S. (1996). Spelling unfamilar names. *Proceedings of the International Congress of Onomastic Sciences.*

Forster, K.I. (1976). Accessing the mental lexicon. In E.C.J. Walker & R.J.Wales (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland.

Foster, K.I. (1981). Priming and the effects of sentence and lexical contexts on naming time: evidence for autonomous lexical processing. *Quartely Journal of Experimental Psychology*, 33A, 465-495.

Foster, K. I., Davies, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation ? *Quartely Journal of Experimental Psychology*, 39, 211-251.

Foster, K.I. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 680-698.

Foster, K.I. (1989). Basic issues in lexical processing. In W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, Mass.: MIT Press.

Foster, K.I., & Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Behaviour*, 12, 627-635.

Foster, K.I., & Taft, M. (1994). Bodies, Antibodies, and Neighborhood-Density effects in masked priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 844-863.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human, Perception and Performance*, 2, 361-379.

Frege, G. (1892) Uber Sinn und Bedeuntung [On sense and meaning]. In G.Patzig (Ed.) *Funktion, Begriff, Bedeutung* (pp 40-65). Gottingen: Vandenhoek und Ruprecht.

Frey, A.W., & Halterman, J.C. (1970). Advertising. New york: Ronald Press.

Friedman, A., & Hall, D. L. (1996). The importance of being upright: Use of environmental and viewer-centered reference frames in shape discriminations of nobel three-dimensional objects. *Memory and Cognition*, 24, 285-295.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, M. Coltheart & J. Marshall (Eds.), Surface Dyslexia, Londom: LEA.
Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, 5, 253-286.

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology*, 74, 159-180.

Gernsbacher, M.A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology:General*, 113, 256-281.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment ? *Psychological Review*, 62, 32-41.

Glushko, R. J. (1981). Principles for pronouncing print: The psychology of phonography. In A.M. Lesgold & C.A. Perfetti, *Processes in reading*. New York: LEA.

Glushko, R.J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 674-691.

Gontijo, P. F. D., Shillcock, R. C. & Kelly, M. L. (1998). Brand names and their contribution to visual word recognition. (submitted).

Goodglass, H., & Butters, N. (1988). Psychobiology of cognitive processes. In R. Atkinson, R. Hermstein, D. Luce, & G. Lindzey (Eds.), *Stevens Handbook of Experimental Psychology, Volume* 2, New York:Wiley-Interscience.

Goodglass, K., & Wingfield, (1993). Dissociations in comprehension of body parts and geographical place names following focal brain lesions. *Memory*, 1, 313-328.

Gough, P. B. (1972). One second of reading. In J.P. Kavanagh & I. G. Mattingly (Eds.) *Language bye Eye and by Ear*. Cambridge, Mass.: MIT Press.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and language*, 29, 228-244.

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception and Psychophysics*, 47, 191-198.

Grainger, J., O'Regan, J.K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition. *Perception and Psychophysics*, 45, 189-195.

Haber, L. R., Haber, R. N., & Furlin, K. R. (1983). Word length and word shape as sources of information in reading. *Reading Research Quartely* 18, 165-189.

Haber, R. N., & Schindler, R. M. (1981). Errors in proof-reading: Evidence of syntactic control of letter processing ? *Journal of Experimental Psychology: Human Perception and Performance* 7, 573-579.

Haber, R. N., and Haber, L.R. (1981). Visual components of reading process. *Visible Language* 15,147-182.

Halle, M. (1990). Phonology. In D. N. Osheron & H. Lasnik (Eds.), *Language: An invitation to cognitive science* (pp. 43068). Cambridge, MA: MIT Press.

Hanna, J., & Hanna, P. (1959). Spelling as school subject: A brief history. *National Elementary Principal*, 38, 8-23.

Harm, M. W., & Seidenberg, M. S. (in press). Phonology, reading acquisition, and dyslexia: insights from connectionist models.

Hart, J., Berndt, R.S., & Caramazza, A. (1985). Category-specific naming deficit folowing cerebral infraction. *Nature*, 316, 439-440.

Henderson, L. (1974). A word-superiority effect without orthographic assistance. *Quartely Journal of Experimental Psychology*, 26, 301-311.

Henderson, L. (1982). *Orthography and word recognition in reading*. London: Academic Press.

Henderson, L., & Chard, M. J. (1976). On the nature of the facilitation of visual comparisons by lexical membership. *Bulletin of the Psychonomic Society*, 7, 432-434.

Howard, D. (1987). Reading without letters ? In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 27-58). London: Erlbaum.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, 41, 401-410.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 148, 574-591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154.

Humphreys, G.W. (1981). Direct versus indirect tests of the information available from masked displays - what visual masking does and does not prevent. *British Journal of Psychology*, 72, 323-330.

Inhoff, A. W & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, 40, 431-439.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception & Performance*, 1, 130-136.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, 29, 687-715.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, 13, 278-305.

Kamin, L.J. (1958). Ethnic and party affiliations of candidates as determinants of voting. *Canadian Journal of Psychology*, 12, 205-212.

Katz, L., & Feldman, L.B. (1981). Linguistic coding in word recognition: Comparisons between a deep and shallow orthography. In A. M. Lesgold & C.A. Perfetti, *Interactive Processing in Reading.* Hillsdale:LEA.

Kay, J. & Bishop, D. (1987). Anatomical differences between nose, palm, foot. Or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp.449-469). Hove, UK:Erlbaum.

Kay, J. (1987). Phonological codes in reading: Assignment of sub-word phonology. In A. Allport, D. G. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language Perception and Production.* : London:Academic Press.

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990), Which elaborative inferences are drawn during reading ? A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp 377-402). Hillsdale, N.J: Erlbaum.

Kripke, S. (1980) *Naming and Necessity*. Oxford: Basil Blackwell.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Laxon, V., Masterson, J., Pool, M., & Keating, C. (1992). Nonword naming: Further exploration of the pseudohomophone efect in terms of orthographic neighborhood size, graphemic changes, spelling-sound consistency, and reader accuracy. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 730-748.

Lettvin, J.Y., Maturana, H. R., McCulloch, W.S., & Pitts, W. H. (1959) What the frog's eyes tells the frog's brain. Proc. Inst. Radio Engr, 47, 1940-1951.

Liberman, I. Y., & Shankweiler, D. (1979). Speech, the alphabet, and teaching to read. In L. B. Resnick & P.A. Weaver (Eds.), *Theory and Practice of Early Reading*. Hillsdale, NJ: LEA.

Lorch, R. F., Balota, D. A., & Stamn, E. G. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre- or post-lexical access. *Memory and Cognition*, 9, 587-598.

Luchelli, F., & De Renzi, E. (1992). Proper name anomia. *Cortex*, 28, 221-230.

Mackay, D. G. (1972). The structure of words and syllables: evidence from errors in speech. *Cognitive Psychology*, 3, 210-227.

Manelis, L. (1974). The effect of meaningfulness in tachistoscopic word perception. *Perception and Psychophysics*, 16, 182-192.

Marshall, J.C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistics approach. *Journal of Psycholinguistics Research*, 2, 175-199.

Masson, M. E. J. (1986). Identification of typographically transformed words: Instance-based skill acquisition. *Journal of Experimental Psychology: Learning Memory and Cognition* 12: 479-488.

Mayall, K., & Humphreys, G. W. (1996). Case mixing and the task sensitive disruption of lexical processing. *Journal of experimental psychology: Learning, Memory and Cognition,* 22, 278-294 .

McCann, R.S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 13-24.

McCannan, R.S., & Besner, D. (1987). Reading pseudohomophones: Implications fro models of pronunciation and the locus of the word-frequency effects in word naming. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 14-24.

McCarthy, R.A., & Warrington, E.K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex*, 22, 359-380.

McClelland, J. L. (1976). Preliminary Letter Identification in the Perception of Words and Nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 80-91.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375-407.

McConkie, G. W. & Zola, D. (1979). Is visual information integrated across successive fixations in reading ? *Perception and Psychophysics*, 25, 221-224.

McKenna, P., & Warrington, E.K. (1978). Category-specific naming preservation: A single case study. *Journal of Neurology, Neurosurgery and Psychiatry*, 41, 571-574.

McKenna, P., & Warrington, E.K. (1980). Testing for nominal dysphasia. *Journal of Neurology, Neurosurgery and Psychiatry*, 43, 781-788.

McWeeny, K.H., Young, A.W., Hay, D.C., & Ellis, A.W. (1987). Putting names to faces. *British Journal of Psychology*, 78, 143-149.

Metchild, R., Markowitsch, H.J., Rauch, M., & Kessler, J. (1997). Differential impairments in recalling people's names: A case study in search of neuroanatomical correlates. *Neuropsychologia*, 35, 667-684.

Millis, M . L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory and Cognition*, 17, 141-147.

Monk, A. F., & Hulme, C. (1983). Errors in proofreading: evidence for the use of word shape in word recognition. *Memory and Cognition* 11, 16-23.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In *Basic Processes in Reading: visual word recognition* (Eds.) D. Besner & G.W. Humphreys, UK.:Erlbaum.

Monsell, S. Doyle, M.C., & Haggard, P.N. (1989). Effects of frequency on visual word recognition tasks. Where are they ? *Journal of Experimental Psychology: General*, 118, 43-71.

Morrison, C. M., Ellis, A.W., & Quinlan, P.T. (1992). Age of acquisition, not word frequency affects object naming, not object recognition. *Memory and Cognition*, 20, 705-714.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165-178.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P.A. Kolers, M. Wrolstad, & H. Bouma (Eds.), *Processing of visible language, Vol 1*. New York: Plenum.

Neely, J. H., Keef, D. E., & Ross, K. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective relation-checking. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 1003-1019.

Neisser, U. (1966). *Cognitive Psychology*. New york, Appleton-Century-Crofts.

O'Sullivan, C.S., Chen, A., Mohapatra, S., Sigelman, L., & Lewis, E. (1988). Voting in ignorance: The politics of smooth-sounding names. *Journal of Applied Social Psychology*, 18, 1094-1106.

Paap, K. R., Newsome, S., McDonald, J.E., & Schvaneveldt, R.W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review*, 89, 573-594.

Paap, K. R., Newsome, S.L., and Noel, R.W. (1984). Word shapes in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance* 10, 412-428.

Patterson, K.E., & Marcel, A.J. (1977). Aphasia, dyslexia and the phonological coding of written words. Quartely Journal of Experimental Psychology, 29, 307-318.

Patterson, K.E., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K.E. Patterson, J.C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 335-359). Hove, UK:Erlbaum.

Plaut, C. D., Seidenber, M.S., McClelland, J. L., & Patterson, K.E. (1994). Understanding normal and impaired word reading: computational principals in quasi-regular domains (Tech. Rep. PDP-CNS-94/5).

Polk, A. T., & Farah, M. J. (1995). The Statistics of the Environment Affect the Functional Architecture of Vision in Adulthood: A Reduced Alphanumeric Category Effect in Canadian Mails Sorters. *Proceedings of the 17th Annual Conference of the Cognitive Science Society.*

Polk, T. A., & Farah, M. J. (1994). Late experience alters vision. *Nature*, 376, 648-649.

Pollatsek, A., Well., A. D. & Schindler., R. M. (1975). Familiarity affects visual processing of words. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 328-338.

Ponansky, C. J. and Rayner, K. (1977). Visual-feature and response components in a picture-word interference task with beginning and skilled readers. *Journal of Experimental Child Psychology*, 24, 440-460.

Posner, M.I., & Mitchell, R. F. (1967). Chronometric Analysis of Classification. *Psychological Review*, 74, 392-409.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, 27, 399-415.

Pring, L. (1981). Phonological codes and functional spelling units: Reality and implications. *Perception and Psychophysics*, 30, 573-578.

Pulvermuller, F. (in print). Words in the brain's language. Behavioural and Brain Sciences, Cambridge University Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J.(1985). *A Comprehensive Grammar of the English Language*. London: Longman.
Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice-Hall International, London.

Reason, J.T., & Lucas, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J.E. Harris & P.E. Morris (Eds.), *Everyday Memory, Actions and Absentmindedness*. London. Academic Press.

Regan, J. (1978). Involutary automatic processing in color naming tasks. *Perception and Psychophysics*, 24, 130-136.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.

Riddoch, J., Humphreys, G., Cleton, P., & Fery, P. (1990). Interaction of attentional and lexical processes in neglect dyslexia. *Cognitive Neuropsychology*, 7, 459-517.

Rubenstein, H., Lewis, S. S., & Rubenstein, M.A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behaviour*, 10, 645-658.

Rubin, D.C., Stoltzfus, E.R., & Wall, K.L. (1991). The abstraction of form in semantic categories. *Memory and Cognition*, 19, 1-7.

Rudnicky, A. I., & Kolers, P. A. (1984). Size and case of type as stimuli in reading. *Journal of Experimental Psychology: Human Perception and Performance* 3, 1-17.

Rumelhart, D. E., & Siple, P. (1974). Process of recognizing tachistoscopically presented words. *Psychological Review* 81, 99-118.

Rumelhart, D.E., & McClelland, J.L., (1986). On learning the past tense of English verbs. In D.E. Rumelhart & J.L. McClelland & the PDP Research Group, *Parallel distributed processing: Vol. 2: Foundations*. Cambridge, MA:MIT Press.

Russell, B. (1905). On denoting. *Mind*, 14, 479-493.

Saffran, E. M. & Marin, O. S. M. (1977). Reading without phonology: Evidence from aphasia. *Quartely Journal of Experimental Psychology*, 29, 515-525.

Schonell F.J., & Brown, G. (1995). A Spelling List for Seniors. Oliver & Boyd, Essex, Longman.

Searle, J.R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14, 489-537.

Seidenberg, M.S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre and post-lexical loci of scripts in reading. *Journal of Memory and Language*, 24, 235-270.

Seidenberg, M. S. (1995). Visual word recognition: an overview. In Miller, J. L. & Eimas, P. D. (Eds.), *Speech, Language, and Comunication*. New York, Academic Press.

Seidenberg, M.S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.

Seidenberg, M.S., & McClelland, J.L. (1990). More words but still no lexicon: Reply to Besner et. al. (1990). *Psychological Review*, 97, 447-452.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *The mechanisation of thought processes*. London: H. M. Stationery Office.

Seymour, P. H. K., & Jack, M. V. (1978). Effects of visual familiarity on "same" and "different" decision processes. *Quartely Journal of Experimental Psychology*, 30, 455-470.

Shallice, T., & Warrington, E.K. (1975). Word recognition in a phonemic dyslexic patient. *Quartely Journal of Experimental Psychology*, 37A, 507-532.

Shankweiller, D. & Liberman, I. (Eds.). (1989). *Phonology and reading disability: Solving the reading puzzle*. Ann Arbor, Michigan: University of Michigan Press

Shillcock, R., & Gontijo, P. D. F. (1995). Capitalisation influences pronunciation:Implications for processing architectures. Poster presented at *AMLap-95 - Architecture and Mechanisms for Language Processing*, Edinburgh,1995.

Shillcock, R.C., & Bard, E.G. (1993). Modularity and the processing of closed class words. In G.T.M. Altmann & R.C. Shillcock, R.C. (Eds.). *Cognitive models of speech processing.* The Second Sperlong Meeting. London:Erlbaum.

Shillcock, R.C., & Kelly, M. L. Neighbourhood effects for function words, but not content words, in the clarity gating paradigm (submitted).

Shillcock., R.C., & Hackett, K. (1995). Intact higher-level constraints on the pronunciation of new written words by nonfluent dysphasics. *Brain and Cognition*, 28, 196-197.

Smenza, C., & Sgaramella, T.M. (1993). Production of proper names: A clinical case study of the effects of phonemic cueing. *Memory*, 1, 265-280.

Smenza, C., & Zettin, M. (1988). Production of proper names: A case of selective inability. *Cognitive Neuropsychology*, 5, 711-721.

Smenza, C., & Zettin, M. (1989). Evidence from aphasia for the role of proper names as pure referring expressions. *Nature*, 342, 678-679.

Smith, F. (1969). Familiarity of configuration vs discriminability of features in the visual identification of words. *Psychonomic Science*, 14, 261 & 263.

Smith, F. (1971). *Understanding Reading*. New York: Holt, Rinehart and Winston.

Springer, S.P., & Deustch, G. (1993). *Left brain, right brain*. NewYork: Freeman Company.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1980). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In R. A. Cole (Ed.), *Perception and Production of fluent speech* (pp. 469-505). Hillsdale, NJ:Erlbaum.

Taft, M. (1991). *Reading and the mental lexicon*. Hove, UK:.Erlbaum.

Tanenhaus, M.K., & Lucas, M.M. (1987). Context effects in lexical processing. In U. Frauenfelder & Tyler (Eds.), *Spoken word recognition*. MIT Press.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, 26, 608-631.

Treisman, R. (1960). Contextual cues in selective listening. *Quartely Journal of Experimental Psychology*, 12, 242-248.

Treisman, R. (1961). *Attention and Speech*. Unpublished doctoral dissertation, Oxford University.

Treisman, R. & Chafetz, J. (1987). Are there onset and rime-like units in printed words? In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, (pp 281-298) Hillsdale, NJ.: Erlbaum.

Treisman, R. (1989). The internal structure of the syllable. In G. Calrson & M. Tanenhaus (Eds.) Linguistic structure in language processing (pp. 27-52). Norwell, MA:. Kluwer Academic.

Treisman, R., & Zukowski, J. L. (1988). Units in reading and spelling. *Journal of Memory and Language*, 27, 466-477.

Treisman, R., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use and acquisition of English orthography. *Journal of Experimental Psychology: General*, 124, 107-136.

Valentine, T., Bredart, S., Lawson, R., & Ward, G. (1991). What's in a name ? Access to information from people's names. *European Journal of Cognitive Psychology*, 3, 147-176.

Valentine, T., Brennen, T., & Bredart, S. (1996). *The cognitive psychology of proper names.* London: Routledge.

Venezky, R.L. (1970). *The structure of English orthography.* Mouton:The Hague.

Voice, J.K.(1995). *Competion Processes in Visual Word Recognition.* Doctoral dissertation. Edinburgh: Edinburgh University. Centre for Cognitive Science.

Warren, R. E. (1977). Time and the spread of activation in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 458-466.

Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology*, 73, 117-129.

Warrington, E. K., & Clegg, F. (1993). Selective preservation of place names in an aphasic patient: A short report. *Memory*, 4, 281-288.

Warrington, E.K., & McCarthy. (1987). Categories of knowledge. Further fractionations and an attempted integration. *Brain*, 110, 1273-1296.

Wernicke, K. (1874). Der aphasische Symptomenkomplex. Breslau [Translated in *Boston Studies in Philosophy of Science*, 4, 34-97.]

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143-154.

Wheeler, D. D. (1970). Processes in Word Recognition. *Cognitive Psychology,* 1, 59-85.

Wickelgren, W.A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 76, 1-15.

Wijik, A. (1966). *Rules of pronunciation for the English language*. Oxford: Oxford University Press.

Young, A. W., Hay, D.C. & Ellis, A.W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology*, 76, 495-523.